University of Massachusetts Amherst
ScholarWorks@UMass Amherst

Doctoral Dissertations | Dissertations and Theses

July 2018

The Neural Correlates of Stereotype Threat and the Stereotype Inoculation Model in Young Women

Chaia Flegenheimer

Follow this and additional works at: https://scholarworks.umass.edu/dissertations_2

Part of the Other Neuroscience and Neurobiology Commons, and the Social Psychology Commons

Recommended Citation
Flegenheimer, Chaia, "The Neural Correlates of Stereotype Threat and the Stereotype Inoculation Model in Young Women" (2018). *Doctoral Dissertations*. 1238.
https://scholarworks.umass.edu/dissertations_2/1238

This Open Access Dissertation is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

The Neural Correlates of Stereotype Threat and the Stereotype Inoculation Model in Young Women

A Dissertation Presented

by

CHAIA M. FLEGENHEIMER

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2018

Neuroscience and Behavior Program

© Copyright by Chaia M. Flegenheimer 2018
All Rights Reserved

The Neural Correlates of Stereotype Threat and the Stereotype Inoculation Model in Young Women

A Dissertation Presented
By
CHAIA FLEGENHEIMER

Approved as to style and content by:

Jennifer McDermott, Chair
Psychological and Brain Sciences Department

Nilanjana Dasgupta, Member
Psychological and Brain Sciences Department

Laurel Smith-Doerr, Member
Sociology Department

Youngbin Kwak, Member
Psychological and Brain Sciences Department

Lisa Sanders, Member
Psychological and Brain Sciences Department

Paul Katz, Interim Graduate Program Director
Neuroscience and Behavior Program

John Lopes, Interim Director
Interdepartmental Programs in Life Science

ACKNOWLEDGMENTS

I would like to thank my advisor, Jennifer M. McDermott, for her many years of guidance and support. Thanks also to Nilanjana (Buju) Dasgupta and Laurel Smith-Doerr, whose contribution to my professional development has been invaluable. I would also like to extend my gratitude to the other members of my committee, Youngbin Kwak and Lisa Sanders, for their helpful comments and suggestions on all stages of this project.

I want to thank the UMass Center for Research on Families for supporting me and helping to fund me through this past year, so that I could spend more time working on these projects.

I am also extremely thankful to all the undergraduate researchers who helped me prepare, run and process these studies. In particular, thank you to Tyler Schreyack, and Jacob Gitlin for helping me do all the frustrating but necessary tasks in the beginning, like entering all 600 images into the study program one at a time. I wish to express my appreciation to all the individuals who participated in this project.

A special thank you to everyone who supported me and helped me stay focused during this process. I would especially like to thank my mother, Jean Flegenheimer, for pushing me to keep going when the going got tough, and my friends and lab-mates, Adaeze Egwuatu, Claudia Lugo-Candelas, Abigail Fontaine, Kathryn Anzuoni and Sarah-Jo Torgrimson, for keeping me grounded and laughing for the last five years.

ABSTRACT

THE NEURAL CORRELATES OF STEREOTYPE THREAT AND THE STEREOTYPE INOCULATION MODEL IN YOUNG WOMEN

MAY 2018

CHAIA FLEGENHEIMER, B.A., VASSAR COLLEGE

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Jennifer McDermott

A promising intervention technique for stereotype threat effects is the stereotype inoculation model (SIM), which utilizes in-group role models to counteract stereotype-induced pressures. However, it remains unclear how the SIM may impact neural mechanisms during stereotype threat, including negative feedback bias (increased attention to undesirable feedback). The following three studies aim to examine the behavioral (Study 1) and neural (Study 2) markers of ST in women and how these markers are influenced by the SIM (Study 3). In each study, participants completed a non-traditional math task (the approximate number task). In the first two studies, one group was told the task was a measure of math intelligence (stereotype threat), while the other group was told it measured creative ability (non-threat). Study 1 focused on the behavioral impact of implicit ST including performance on the task, as well as motivation to continue with the task, and confidence within the task. Men and women were both included as participants. ST negatively impacted motivation to continue the task in women, but not men. In addition, higher math identification related to lower immediate task performance, but higher task confidence and motivation for women in the ST condition. Study 2 explored the neural mechanisms underlying implicit ST in women,

particularly focusing on performance monitoring measured using event-related potentials (ERPs) to assess performance processing. Waveforms associated with internal response-monitoring were negatively impacted by ST as evidenced by inefficient response-monitoring and more conscious focus on errors. In Study 3, all participants were told the task measured math ability, and groups were given difference biographies to read prior to task completions. The biography conditions were 1) consistent with stereotype threat (male mathematicians), 2) the SIM (female mathematicians) and 3) a non-threat collection (mixed-gender artists). The SIM condition impacted the participants' perception of the task, such that anxious women viewed it as more of a game, whereas participants in the ST condition perceived the task as a test. Women in the SIM condition also exhibited greater neural reactivity to correct responses prior to the onset of external feedback, and less overall neural reactivity to external feedback cues.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

BACKGROUND: STEREOTYPE THREAT AND THE SIM

1.1 Gender gap in STEM fields

Women remain underrepresented in science, technology, engineering, and math (STEM) fields, with the percentage of women decreasing as the positions become more advanced (NSF, 2015; Beede et al., 2011). In 2014 the National Science Foundation (NSF) reported that women received fewer than twenty-five percent of undergraduate and graduate degrees in computer science and engineering (2015), a statistic that remained relatively static for at least five years (Beede et al., 2011). Rather than the number of women increasing in recent years, there seems to have been a decrease in the number of women receiving these STEM degrees, with recent reports showing women earning only eighteen percent of the computer science and engineering degrees (Ashcraft, Eger & Friend, 2012). There is also a significantly larger attrition rate for women in science graduate programs compared to men, such that women who enter these programs are more likely to leave prior to program completion (Ferreira, 2003). Moreover, only about twenty-six percent of women with college degrees in STEM subjects go on to work in STEM jobs (Beede et al., 2011). Furthermore, women in STEM jobs have higher turnover intentions compared to their male peers (Cech, Rubineau, Silbey & Seron, 2011; Xu, 2007), and are more likely to leave their field of study compared to women in other professional occupations (Glass, Sassler, Levitte & Michelmore, 2014).

Early explanations for this gender gap in the STEM fields suggested that women were less capable of doing STEM tasks compared to men, and therefore failed in STEM-related jobs (Benbow & Stanley, 1982; Shields, 1975). More recent studies contradict this initial theory and instead support the idea that there is no inherent gender difference in STEM ability. For instance, throughout grade school and high school, girls' math and science performance matches or exceeds that of their male peers (Hyde, Lindberg, Linn, & Williams, 2008; Lindberg, Hyde, Petersen & Linn, 2010; Voyer, & Voyer, 2014). This trend follows through to college math courses, where women earn equal or slightly higher mathematics grades compared to men (Bridgeman & Wendler, 1991; Benbow & Stanley, 1982). Although men tend to perform better than women on math-related standardized tests, this gap in performance disappears in countries with higher levels of gender equality, suggesting that the gaps are socially driven rather than actual differences in ability (Guiso, Monte, Sapienza, & Zingales, 2008; Hyde & Mertz, 2009). Overall, this research suggests that women are just as able to perform well in STEM subjects as men. However, despite equal capability, women continue to stray away from higher degrees and jobs in STEM fields.

The persistent gender gap in STEM fields is troubling for several reasons. First, it lessens the overall diversity of scientific perspectives; what are chosen as important areas of study and viable methods of research are often based on researchers' personal experiences and interests (Medin & Lee, 2012). Diversity among researchers is needed to push scientific study to new and innovative places, and to answer questions that impact diverse populations. Second, the continued gender gap in STEM fields may also limit the number of higher paying jobs for women. Women in the STEM fields earn an average of 33% more money compared to their non-STEM counterparts (Beede et al. 2011). Furthermore, the gender-based wage gap is smaller in STEM occupations compared to non-STEM occupations, with women earning 14 percent less than men in STEM fields compared to a 21 percent difference in non-STEM fields (Beede et al., 2011).

Beyond better science through research diversity and higher-paying job opportunities for women, closing the gender gap in STEM fields is also important for the American economy. STEM jobs are an essential part of the current US economy, and the need to fill STEM jobs is outpacing the growth rate of non-STEM jobs (Pham & Triantis, 2015). By the year 2022, approximately 6.6 million STEM jobs need to be filled for the US to remain competitive in the global economy. These jobs could be filled by international employees, however, that path does not help the general American work force, and makes American scientific innovative progress more dependent on the international political climate (Branch & Alegria, 2016). As women make up half of the American workforce, it is unlikely that these jobs can be filled by American workers without a large increase in female scientists, engineers and mathematicians (Pham & Triantis, 2015; Dasgupta & Stout, 2014). Therefore, closing the gender gap in STEM fields is beneficial for scientific discovery, women and to the continued growth of the American economy. With research showing that this gender gap is not driven by differences in ability, the question then becomes, 1) what is stopping women from joining STEM fields at the same rate as men and 2) how can this pattern be counteracted?

1.2 Stereotype threat

More recent explanations for the gender gap in STEM fields suggest that women stray away from positions, particularly high-powered positions, in STEM fields due to a lack of motivation and interest in the natural sciences (Wang, Eccles, & Kenny, 2013; Sadler, Sonnert, Hazari, & Tai, 2012; Ferriman, Lubinski, & Benbow, 2009; McArdle, 2008; Rosenbloom, Ash, Dupont, & Coder, 2008). Whereas motivation undoubtedly plays a role in career choice, this explanation often ignores the influence of implicit pressures and negative stereotypes, which consistently indicate to women that they will not succeed and do not belong in STEM fields (Cheryan, 2012; Shapiro, & Williams, 2012; Galdi, Cadinu, & Tomasetto, 2014; Deemer, Thoman, Chase & Smith, 2014).

Although there are instances of explicitly stated negative gender stereotypes within the STEM fields (i.e. Summers, 2005), stereotyped ideas are commonly expressed in subtle and often implicit ways, starting from a very early age. Even as early as elementary school, teachers have reported lower expectations for math achievement among their female students compared to their male students (Mizala, Martinez & Martinez, 2015). These differences in expectations likely influence teachers' behavior towards their male and female students, as research has shown that teachers provide more praise for successes performed by students they place higher expectations on, and are less likely to spend time with students they have low expectations for (Babad, 2009; Jackson & Leffingwell, 1999; Brophy & Good, 1970). Parents also often contribute to implicit ST effects in subtle ways. For example, parents are more likely to attribute girls' math-related successes to effort and boys' math-related successes to innate talent, suggesting that girls need to work harder than boys to achieve equivalent results in math-related fields (Gunderson, Ramirez, Levine & Beilock, 2012; Yee & Eccles, 1988; Raty, Vanska, Kasanen & Karkkainen, 2002; Tiedemann, 2000). Later, in higher levels of education, there are also a growing number of implicit environmental cues. For instance, the ratio of male to female students within STEM fields becomes increasingly unbalanced at higher levels of education. Male students far outnumber female students in STEM fields starting in high school and becoming more drastic at each level of education, emphasizing the supposed discrepancy in male and female STEM capabilities (Ceci & Williams, 2010; Ibarra, Carter & Silva, 2010). By the time women are young adults, they have been exposed to years of subtle and implicit gender sterotype cues indicating that they will not do well in STEM fields.

These gendered cues, perpetuating overall negative stereotypes can impair women's performance on STEM related tasks by creating an environment in which women feel the need to prove their competency, either as a positive representative of the female gender or as an exception to the rule, a phenomenon known as Stereotype Threat (ST) (Aronson, Quinn & Spencer, 1998; Dasgupta, 2011). Interestingly, ST is not reliant on women believing that they conform to the negative stereotypes, but rather on their desire to disprove them. In fact, ST seems to have the greatest impact on individuals who highly identify with the stereotyped domain (Aronson, Lustina, Good, Keough, Steele & Brown, 1999; Keller, 2007). For example, Keller found that girls who highly identified with math performed worse on difficult math questions when under ST, whereas low math-identifying girls performed better on difficult questions when they were under threat compared to non-threat (2007). Keller's finding suggests that women who do not identify with the threatening field, who have nothing to lose, may be able to rise to the challenge that ST poses, whereas their high-identifying peers may be hindered by the pressure to show their capabilities. Supporting this interpretation, ST has been linked with both increased anxiety and vigilance, as measured by physiological techniques such as skin conductance and heart rate (Murphy, Steele & Gross, 2007; Osborne, 2007). In response to increased stress, women under ST have demonstrated increased efforts towards emotional regulation (Johns, Inzlicht & Schmader, 2008) and suppression of negative thoughts (Schmader, Johns & Forbes, 2008), which takes resources away from task-critical abilities. Even in young girls, ST induces a decreased ability to complete difficult math problems, but increased motivation to complete easy math questions, suggesting lower levels of mental resources to complete the more complicated problems (Neuville & Croizet, 2007). Indeed, women under ST show lower executive functioning ability, including lower inhibition and updating (Rydell, Van Loo, & Boucher, 2013), which impairs women's performance on cognitive tasks (Schmader et al., 2008).

Beyond performance deficits, chronic ST is thought to lead targeted individuals to disengage from the stereotyped domain (Crocker & Major, 1989; Beasley & Fischer, 2012; Steele, 1997). As stated earlier, ST affects individuals who highly identify with the threatened field, meaning that their sense of self-worth is related to their capabilities within the negatively stereotyped domain. Disengagement from stereotyped tasks is believed to serve as a coping mechanism to protect aspects of self-identity such as self-esteem (Crocker & Major, 1989; Major, Spencer, Schmader, Wolfe & Crocker, 1998; Woodcock, Hemandez, Estrada & Schultz, 2012). By psychologically distancing themselves from the outcome of a stereotyped task people are shielding their self-confidence from perceived setbacks, increasing their ability to endure within a threatening field. Indeed, in the short-term, disengagement with stereotyped tasks increases the likelihood that threatened individuals will persevere through a threatening task (Nussbaum & Steele, 2007).

Although protective in the short-term, disengagement can lead to overall de-identification with the stereotyped field. De-identification encompasses individuals' view that the stereotyped field is not important for their future success, lowering their motivation to continue and succeed within that field. There is growing evidence that suggests that ST leads to lower levels of engagement and motivation within a stereotyped task. For instance, women's self-esteem seems to be less affected by feedback valence on a math task than men's self-esteem, indicating that women under ST are disengaging their sense of self-worth from their task performance (Fogliati & Bussey, 2013). Furthermore, women experiencing ST have shown less willingness to use a math tutor voluntarily after receiving negative feedback during a math task, signaling their disengagement from the task and subsequent lower drive to succeed (Fogliati & Bussey, 2013; Mangels, Good, Whiteman, Mapiscalo & Dweck, 2011). Overall, these results support the use of protective disengagement and the potential for de-identification by female participants during a stereotyped task.

Despite the negative potential of de-identification for overall motivation in threatening fields, the role of high domain identification is less clear. In the ST literature, high domain identification is often viewed as a vulnerability. For instance, women with high math identity have demonstrated lower task performance under threat, particularly on difficult task items (Keller, 2007). Similarly, African-American students who have higher academic identity ratings have higher rates of school withdrawal (Osborne & Walker, 2006). Combined, these findings suggest that individuals with higher domain identity are more vulnerable to the increased stress associated with ST, as evidenced by immediate performance difficulties and long-term dis- engagement from threatening tasks. However, students with high academic identity also demonstrate higher GPAs and fewer school absences, regardless of race (Osborne & Walker, 2006), indicating increased task engagement with higher domain identification. Indeed, domain identification is generally positively related to intrinsic motivation and meaningful cognitive engagement (Walker, Greene & Mansell, 2006), suggesting a potential protective factor against ST effects on task motivation, which could be key to future ST interventions. The possible protective effect of domain identification within ST will be explored in Study 1.

1.3 Stereotype threat and neural processing

As many of the ST cues are subtle and implicit, many of the effects within individuals follow the same patterns. For example, people often explicitly deny having prejudices, while implicitly acting on them (Greenwald & Banaji, 1995; Greenwald et al., 2002). Indeed, people's implicit associations are thought to mediate the impact of gender-based stereotype threat cues (Galdi, Cadinu & Tomasetto, 2013), whereas explicit opinions on gender roles have little effect (Huguet & Regner, 2009). Over the years, social psychology has come to rely less on explicit data, such as self-report, and more on implicit measures, like the Implicit Attitudes Test, to gain a more unbiased understanding of social phenomenon (Derks, Inzlicht & Kang, 2008). Towards this end, neuroscience methods have begun to be incorporated with social psychological concepts through a developing field coined social neuroscience (Cacioppo, Berntson & Decety, 2010; Amodio, 2010). These techniques can help advance the understanding of social psychological processes by providing more knowledge of the neural mechanisms underlying social behavior as well as an established method for measuring implicit reactions.

Early work using social neuroscience to explore stereotypes and prejudices focused on understanding the perpetrators to better grasp the mechanisms underlying implicit bigotry (Wheeler & Fiske, 2005; Phelps et al., 2000; Hart et al., 2000). More recently, research has begun to explore the effect of stereotypes on the neural processing of the stigmatized groups (e.g. Wraga et al., 2006; Forbes & Leitner, 2014). Two major neuroscience techniques that have been used in this research are functional magnetic resonance imaging (fMRI) and event-related potentials (ERPs).

1.3.1 fMRI. Using fMRI methods allows researchers to non-invasively link behaviors with enhanced or decreased activation in specific brain regions. By exploring the effects of ST on multiple neural regions at once, fMRI data can provide information about multi-pronged mechanisms underlying ST effects. This information can then be used to supplement behavioral studies to build a complex mechanistic understanding of ST.

To date two studies have explored ST effects using fMRI techniques. The first such study was performed by Wraga, Duncan, Jacobs, Helt and Church (2006) and explored the role of negative and positive stereotypes on women's neural activation and performance on a mental-rotation task. Young women were exposed to either a negative stereotype (i.e. women perform worse on spatial reasoning tasks), a positive stereotype (i.e. women are better at adapting perspectives), or neutral information and then asked to perform a mental-rotation task. Women in the negative stereotype group performed the worst on the rotation task and showed increased activation in neural regions associated with emotional self-regulation (i.e. rostral-ventral anterior cingulate cortex) and social processing (i.e. right orbital gyrus). In contrast, women in the positive stereotype group showed increased task performance, and activation in task-related brain areas, including those associated with visual processing (i.e. Brodmann areas 18/19) and working memory (i.e. ventral anterior prefrontal cortex). These findings suggest that women exposed to a negative stereotype experience both increased emotional load and decreased executive functioning ability.

A similar study performed by Krendl and colleagues explored the neural basis of women's underperformance in math when confronted by a negative stereotype among young women who reported high math identification (2006). All participants completed a test of baseline math ability prior to the stereotype manipulation. Women in the ST group were then reminded of the negative stereotype that women perform worse in math tasks. Following task manipulation, participants completed a second round of math testing. Women in the control group showed increased math performance over time, along with increased activation in brain areas associated with math learning and performance (i.e. the inferior prefrontal cortex, the inferior parietal cortex and the bilateral angular gyrus). In contrast, women in the ST condition did not exhibit increased activity in these brain regions, and instead showed increased activation in areas associated with emotion and social processing (i.e. the ventral anterior cingulate cortex), as well as a sight decrease in math performance over time. Combined, the results from these two fMRI studies demonstrate that ST decreases task performance in the stigmatized group by inhibiting task-related cognitive processing and simultaneously increasing engagement of regions involved in emotional control.

1.3.2 ERP. Whereas fMRIs are useful for understanding which specific brain regions are involved in ST, ERPs allow researchers to determine *when* stigmatized individuals are demonstrating more or less neural activity during a stereotyped task. The excellent temporal resolution of ERPs can provide information on the timing and dynamic interplay of neural activity underlying ST effects. For example, it seems likely that negative bias is involved in ST effects. Women under ST report higher levels of negative thoughts about their ability and the given task compared to women in a non-threat condition (Cadinu et al., 2005). Furthermore, related anxieties, such as social phobia, involve a negative-bias towards evaluative feedback, including a dismissal of positive feedback (Weeks, Heimberg, Rodebaugh, & Norton, 2008), an overemphasis of negative feedback directed at the self (Brozovich & Heimberg, 2008; Morgan & Banerjee, 2008; Cody & Teachman, 2010), and a negative interpretation of ambiguous feedback (Kashdan & Roberts, 2007). This negative bias in feedback processing occurs too quickly to be captured by fMRI techniques, which take about 5-6 seconds to register a change in neural activity (Haan, & Thomas, 2002). ERPs, which measure brain activity on a scale of milliseconds, can help establish whether this type of quick negative bias in performance monitoring occurs in ST, and if so, how it relates to ST outcomes such as task disengagement.

1.3.2.1 Error-related negativity (ERN) and feedback-related negativity (FRN). ERPs often explored in relation to performance monitoring are the error-related negativity (ERN; and the related correct response negativity [CRN]) and the feedback-related negativity (FRN). Early error-detection during internal error-processing is indicated by the ERN, which is a negative going deflection occurring approximately 50-130ms after a participant's response (Forbes, Schmader & Allen, 2008; Clayson, Clawson & Larson, 2011). The FRN, in contrast, indicates processing of externally given feedback. The FRN is a negative deflection between 250-300ms post feedback which is thought to measure relatively automatic feedback processing.

Both the ERN and the FRN are involved in different aspects of performance monitoring, but seem to be involved in the same system. A combination of source localization studies, which estimate the neural origins of ERPs, and joint fMRI-ERP studies indicate that both the ERN and the FRN are generated by the anterior cingulate cortex (ACC; Ladouceur, Dahl, Birmaher, Axelson & Ryan, 2006; Hauser et al., 2014) which is heavily involved in, among other things, processing information saliency and conflict monitoring (Botvinick, Cohen & Carter, 2004). Moreover, a study by Heldmann, Russeler and Munte showed that the ERN and the FRN activation are interrelated (2008). Specifically, the ERN is activated when the information during a task was sufficient to determine an error without external feedback. When this criterion was met, and an ERN was triggered, then external feedback became redundant and the FRN was attenuated. Similarly, when it was not possible to determine accuracy internally an ERN was not produced, and instead an FRN was created in response to the external feedback information. A more recent study by Stahl confirms this ERN-FRN relation (2010). Stahl had participants complete a task in which errors could be made by pressing the wrong button, which the participants could monitor themselves, or by answering a question too slowly, which required external feedback for reliable error detection. She found that button-press errors were followed by increased ERN but not FRN amplitudes. Also, when reaction time errors were made they were followed by an increase in the FRN amplitude, but not in the ERN amplitude. Overall, there seems to be a feedback mechanism, in which the ERN is activated during tasks that are transparent enough that external feedback is not necessary to accurately monitor performance, and this activation then inhibits the generation of the FRN. Alternatively, if the task is complex enough that external feedback is required for performance monitoring then the ERN is attenuated and the FRN is amplified. However, this relation was shown in healthy students who were not under threat. No studies thus far have explored the association between the ERN and the FRN under ST conditions.

In fact, only a few studies to date have explored the ERN and the FRN in relation to ST. A recent study by Forbes and Leitner measured the FRN in young women while they solved simple multiplication problems (2014). Each math problem was accompanied by a set of three possible answers which participants needed to choose between. Participants were told the task measured either their math intelligence (ST condition) or their problem-solving strategies (NT condition). Participants were given 16 seconds to solve each problem, after which they were presented with feedback in the form of a blue "correct" or a red "wrong". On average participants took 10.69 seconds to respond to each problem. Forbes and Leitner found no group difference in FRN reactivity. However, this study paradigm involves task trials with predictable answers, and a long enough response window to preclude time-based errors, which together suggest that the participants could have been monitoring their performance internally rather than relying on the external feedback. Therefore, the group differences may have been occurring at the earlier ERN rather than the FRN. However, as this study did not measure the ERN it is not possible to determine whether there is truly no difference in FRN-amplitude during ST, or whether this difference was assuaged by an earlier difference in ERN-amplitude.

A similar study by Mangels and colleagues further explored the relation between the FRN and ST (2012). Young women completed two sets of math problems after being told the study was assessing either gender differences in math intelligence (ST) or non-gendered effortful problem-solving (non-threat). Participants had 1 minute to answer each question, and feedback was given as a high tone and green asterisk for correct responses and a low tone and red asterisk for incorrect responses. During the first testing session participants were given the opportunity to explore a tutorial after each question. The extent to which participants used the tutorial was used as a measure of learning engagement. Participants returned the following day and were given a surprise re-test of the math problems without the option of a tutorial. Improvement on this second math-problem session was used as a measure of learning from the previous session. ERPs were recorded only during the first session. While Mangels and colleagues found no overall group difference in FRN amplitudes, they did find a different relation between FRN amplitudes and task engagement in the ST group. This relation was such that women under ST were more likely to disengage from the stereotyped task, as evidenced by less tutor exploration, if they demonstrated larger FRN amplitudes, reflecting heightened feedback processing. No such relation was seen for women in the non-threat group. These findings indicate that negative feedback impacts women under threat in a unique way, such that they are more likely to disengage from the task (less tutor exploration) than women in the non-threat group.

These conclusions are supported, and further expanded on, by Forbes, Schmader and Allen (2008) in an ERP study in Latino and African American students under ST. Forbes and colleagues had minority students first complete a baseline round of a flanker task. Participants were then told that the task was either a pattern recognition task (non-threat) or diagnostic of intelligence (ST) prior to completing a demographic questionnaire, where they indicated their race. They then completed a second round of the flanker task. Incorrect answers were followed by negative feedback, but no positive feedback was given. Forbes and colleagues (2008) found that participants in the ST group showed a correlation between subjective valuing of academics and ERN amplitude, such that valuing academics predicted larger ERN amplitudes (more error monitoring). Furthermore, within the ST group valuing academics also predicted faster reaction times following errors as well as fewer errors overall. These findings suggest that the degree to which stigmatized groups focus on errors is related to their initial level of task identification, such that more task identification predicts more attention to, and more efficient processing of errors under ST.

Overall these studies suggest a unique relation between performance monitoring and behavior under ST that corresponds with the possible relation between ST and domain identification. Specifically, individuals under ST seem to perform more internal error monitoring when the task is important for their sense of self-worth, which also correlates with higher task performance. However, heightened negative feedback processing also relates to more task disengagement for individuals under ST. These findings so far support the two-pronged impact of domain identification under ST, with both protective aspects of heightened error processing with heightened domain identification (better immediate task performance), and longer-term damaging effects (more task disengagement with heightened feedback processing). However, with so few ERP studies exploring ST effects, there are still several gaps to explore. These gaps include the relation between internal error monitoring (ERN) and external feedback processing (FRN) under ST, and the relation between domain identification and FRN amplitude under ST.

1.3.2.2 Post-error positivity (Pe) and the P300. In addition to understanding these relatively early ERPs, exploring ERPs associated with later aspects of performance-monitoring will also be important for elaborating on the mechanisms underlying ST and the best ways to counteract their negative effects. One such ERP is the post-error positivity (Pe), which is a positive deflection 100-200ms after an error, and is theorized to indicate later, more subjective, internal error processing (Nieuwenhuis et al., 2001; Hajcak, McDonald & Simons, 2004; Hughes & Yeung, 2011; Orr & Carrasco, 2011). For example, heightened Pe amplitudes are associated with perceived, but not unperceived errors (Nieuwenhuis et al., 2001), and an increase in post-error slowing, a behavioral mechanisms used to increase the likelihood of correct responses on subsequent trials (Hajcak et al., 2004).

A similar later error-processing ERP is the P300, which is measured as a positive deflection which peaks approximately 300-500ms after an event. Like the Pe, the P300 is thought to indicate more conscious processing of a significant event, linking to a behavioral change (Ridderinkhof, Ramautar & Wijnen, 2009). For example, in a probabilistic reversal learning task, the P300 was larger after expected errors that preceded a behavioral adjustment and smaller following unexpected errors which did not predict behavioral changes (Chase et al., 2011). Overall, the Pe and the P300 are thought the indicate similar top-down performance monitoring related to subsequent behavioral changes, but while the Pe follows a response, the P300 is found after external feedback is given.

Interestingly, the more immediate internal error-monitoring of the ERN may have an inverse relation to social stress compared to the FRN and the later error-processing seen with the Pe. Specifically, from late childhood to adulthood, individuals with high levels of anxious apprehension show increased ERN amplitudes (Moser, Moran & Jendrusina, 2012; Hajcak, McDonald & Simons, 2004). However, these same individuals show smaller Pe amplitudes. This relation could indicate heightened error-monitoring, combined with a decrease in more conscious error-processing related to behavioral changes. Furthermore, in contrast to the ERN, the FRN is shown to be attenuated in anxious individuals during negative feedback, possibly indicating lower outcome surprise during negative results (Gu, Huang & Luo, 2010). This combination of effects could contribute to skewed error-reporting in social anxieties (Brozovich & Heimberg, 2008; Morgan & Banerjee, 2008; Cody & Teachman, 2010; Kashdan & Roberts, 2007), and similar increased negative thinking under social stressors such as ST (Cadinu et al., 2005), along with less use of error-monitoring to improve behavior. However, these results are not consistent across the literature. For example, other studies have shown no relation between the Pe and anxiety (Weinberg, Olvet & Hajcak, 2010). Furthermore, while the P300 has an attenuated amplitude in socially anxious individuals during an oddball task (Sachs et al., 2004), no research that we are aware of has explored this ERP in such patients in response to task feedback. Understanding the relation between the ERN, the FRN and the Pe in ST may help explain related negative biases in error-reporting, and whether they follow the same pattern as that seen in clinically anxious populations.

Overall, understanding the neural patterns of error processing in ST is important for better parsing out the implicit mechanisms underlying ST effects, as well as how to counteract these effects with future intervention work. These routes will be explored in studies 2 and 3 respectively.

1.4 Stereotype interventions and the inoculation model

The Stereotype Inoculation Model (SIM) aims to subtly counteract negative stereotypes while simultaneously allowing individuals to embrace their self-identity. The SIM works by implicitly disproving negative group stereotypes through in-group role models (Dasgupta, 2011; Stout, Dasgupta, Hunsinger & McManus, 2011). This model is centered on the notion that for a stigmatized individual to benefit from in-group role-models they must be able to relate to them (Asgari, Dasgupta & Stout, 2012). Thus, in-group experts who are portrayed as special or unusually gifted will not counteract implicit attitudes within stigmatized groups. It must be clear that the counter-stereotypic role-models are not exceptions to the rule, but rather evidence that the group stereotype is false. The SIM is thought to work as a “social vaccine” by “inoculating” an individual’s sense of self against negative stereotypes, such that their feelings of belonging and self-efficacy are stronger, and they are thus less threatened (Dasgupta, 2011; Dasgupta & Stout, 2014). This type of inclusive approach to stereotype intervention is an improvement over other intervention techniques that have focused on actively distancing individuals from negative group stereotypes.

For instance, one alternative intervention technique focused on blurring intergroup boundaries to lessen group-based stereotype effects (Rosenthal & Crisp, 2006). As this intervention involves the stigmatized group explicitly comparing themselves to the non-stigmatized group, this method could easily slant away from empowering the disenfranchised and move towards alienating them. Indeed, individuals who highly identify with their in-group are more likely to self-stereotype when the similarities between their in-group and out-group are highlighted (Spears, Doosje & Ellemers, 1997). Similarly, another intervention technique had participants focus their attention on their individual traits, taking attention away from their group affiliation (Ambady, Paik, Steele, Owen-Smith & Mitchell, 2004). As group identity can be implicitly activated (Devos & Banaji, 2003), it is not always possible to mentally separate from one's group association. Furthermore, an intervention that focuses on having people separate themselves from a stigmatized group is not ideal for long-term methods, as it does not address the core issue of the prejudice. Overall, this type of intervention technique that focuses on consciously differentiating oneself from the negatively stereotyped group and related group characteristics, is neither ideal in the long-run nor truly applicable in a real-world setting. Therefore, research should focus on self-empowering methods of ST intervention, such as the SIM, which do not force people to either compare themselves to other groups, or distance themselves from their own in-group.

Indeed, there is strong behavioral evidence indicating that the SIM intervention is effective at bolstering positive implicit attitudes (McIntyre et al., 2005; Marx & Roman, 2002; Dasgupta & Asgari, 2004; Stout, Dasgupta, Hunsinger & McManus, 2011). Recently, a series of studies by Stout and colleagues (2011) demonstrated that exposure to in-group experts significantly increased women's implicit attitudes towards STEM, their expended effort on a math task, and their perceived ability within STEM fields. For instance, women who read biographies of female engineers showed more positive implicit attitudes towards math (Stout et al., 2011, study 2). Furthermore, self-reported identification with the female engineers predicted greater self-confidence in personal engineering ability compared to women who read about male engineers. A similar study done by Good and colleagues (2010) showed that female high-school students could better comprehend and retain information from a chemistry textbook when the excerpt was accompanied by an image of a female chemist, compared to a male chemist. Overall, these findings demonstrate that exposure to in-group experts increase individuals' sense of belonging and confidence within the threatening field, and in turn bolster their task performance. Although it may not always be possible to expose women and girls to female role-models in person, increasing the chances of reading about female role-models in school is highly applicable. Therefore, the SIM holds great promise for lessening the gender gap in STEM fields.

What remains to be understood is whether the SIM alters implicit processes of attention and feedback processing as measured by neural markers. For example, the impact of the SIM on performance-monitoring processes may have important implications for long-term effects on motivation to enter and continue within threatening fields. Under ST, as discussed above, individuals become more reactive to errors (Forbes, Schmader & Allen, 2008; Mangels et al., 2008). As each error can become a confirmation of previously held negative stereotypes, increased performance-monitoring (i.e. via heightened attention to errors and negative feedback) may serve to strengthen negative attitudes towards stereotyped fields. Specifically, individuals under ST may enter a task with low self-confidence (Stout et al., 2011), and attend more to the negative feedback given during the task as evidence for their low task ability. Over time, a continued negative bias in performance-monitoring may then contribute to lowering their motivation to continue within the stereotyped field (Mangels et al., 2008). Therefore, if the SIM can protect against increased performance-monitoring under ST, this model could potentially help break the link between stereotypes and loss of motivation among threatened groups, thus helping to increase women's long-term motivation to pursue STEM careers.

1.5 The present study

The present study aims to examine the behavioral and neural effects of implicit ST on task performance and engagement in young adults, as well as the protective impact of the Stereotype Inoculation Model via the following three aims:

Aim 1: **Confirm ST effects** for young women using a non-traditional math task.

- When women are told a task is a measure of mathematical intelligence (ST condition), it is hypothesized that they will show less motivation to continue the task, report less confidence in their ability to do well on the task, and perform worse on the task itself compared to women who are told the task measures creative ability (Non-Threat/NT condition) and men in both of the conditions.
- Higher math identity is predicted to produce contrasting effects for women in the ST group, such that they will demonstrate lower immediate task performance, but increased task motivation and confidence.

Aim 2: Examine the **impact of ST** on young women's reactivity to errors using a non-traditional math task.

- We hypothesize that women in the ST group will demonstrate enhanced error detection (larger error-related negativity/ERN) during easy trials, when internal monitoring is plausible, and enhanced feedback processing (attenuated FRN) during difficult trials, when internal monitoring is more difficult. We also hypothesize the participants will show less prolonged error processing (attenuated

P300 and error positivity/Pe) after being told the task measures math ability (ST group) compared to the NT group. This pattern of error processing would account for increased error saliency during ST conditions, combined with a difficulty to effectively use error information to improve task performance.

- Math identity is predicted to relate to larger ERN amplitudes during easy trials and larger FRN amplitudes to negative feedback during difficult trials in women in the ST group but not in women in the NT group.

Aim 3: Examine whether the **Stereotype Inoculation Model (SIM) protects** against ST effects in young women during a non-traditional math test.

- We hypothesize that women who read about female STEM experts (stereotype inoculation group) will attribute less saliency to their errors as evidenced by an attenuated ERN response during easy trials and an attenuated FRN response to negative feedback during difficult trials. Further, we hypothesize that women exposed to relatable in-group experts will demonstrate increased error processing (increased P300 and Pe) as compared to the ST-consistent group. This pattern suggests the effective use of error information, without enhanced negative error-reactions.

CHAPTER 2

STUDY 1: BEHAVIORAL MEASURES OF ST

2.1 Aims and Hypothesis

The aim of Study 1 was to measure ST effects for young women using a non-traditional math task. We hypothesized that when women were told a task was a measure of mathematical intelligence (ST condition), that they would show less motivation to continue the task, report less confidence in their ability to do well on the task, and perform worse on the task itself compared to women who were told the task measures creative ability (Non-Threat/NT condition) and men in both of the conditions. We further hypothesized that higher math identity would produce contrasting effects for women in the ST group, such that they would demonstrate lower immediate task performance, but increased task motivation and confidence.

2.2 Methods

2.2.1 Participants. One hundred and sixty-four college students were recruited from the University of Massachusetts (UMass) Amherst, including 95 females and 69 males. Of these participants five were excluded due to equipment failure, thirty-five were excluded because of experimenter error, and two were excluded from analysis because they were at least 2.5 standard deviations away from the group math identification mean. Overall, excluded participants did not differ from included participants on age, $t(160) = -.119$ $p=.906$, general anxiety, $t(160) = .414$ $p=.680$, general stress, $t(160) = .011$ $p=.991$, math identification score, $t(160) = -1.208$ $p=.229$, or distribution amongst the two conditions, $X^2(1) = .284$, $p = 0.594$, $\varphi=-.042$. The final sample of one-hundred and twenty-two included 62 females ($M= 19.7$ years, $SD = 1.3$) and 60 males ($M= 19.6$ years, $SD = 1.1$).

Final sample size was determined based on previous studies exploring motivation under stereotype threat in college-age women (Fogliati & Bussey, 2013; Forbes & Schmader, 2010). The final sample size was confirmed using data from 16 female pilot participants, who were run using a preliminary version of the numerical discrimination task. Pilot results for the condition-by-fourth-round-decision indicated large effect sizes ($X^2(1) = 6.349$, $p = 0.012$, $\varphi=0.630$), and power analysis indicated that approximately 30 participants per group would be needed to achieve 0.9 power with alpha set at 0.01. To participate, students needed to be at least 18 years of age, and they needed to have completed the pre-screen survey via the UMass Amherst participant recruitment SONA System. Students could not participate if they had been diagnosed with a learning or attention disability, or if they were colorblind. Students were compensated for their time with extra credit points in participating psychology courses.

2.2.2 Procedure. Participants were randomly assigned to either the stereotype threat (ST) condition or the non-threat (NT) condition. Groups of up to four female or male participants were brought into the lab at a time; groups were either all female or all male at a time. Participants were greeted by a male research assistant and told that they would be performing either a measure of math intelligence (ST condition) or creative ability (NT condition) to better understand how females (told to female participants) or males (told to male participants) processed visual learning cues. Based on methods used previously (Stout, Dasgupta, Hunsinger, & McManus, 2011), the two experimental groups were further distinguished by the shirts the research assistant wore, with experimenters wearing either a t-shirt with the quadratic equation (ST condition) or an artistic depiction of the sun (NT condition). Participants were asked to rate their predicted ability on the task (described in more detail below). Next participants completed the practice blocks for the numerical discrimination task. Each trial began with a fixation mark (250ms), which was followed by a dot image that was displayed for 200 ms with a total response window of 1000ms. After a response, accuracy feedback was presented for 1000ms. Feedback consisted of a "thumbs up" for correct answers or a "thumbs down" for incorrect answers. The task speed increased across the first three blocks of practice, such that in the first block the dot images were shown for 700ms, in the second block they were shown for 500ms, and in the third block they were shown for 200ms ('game speed'). After the initial practice blocks participants were offered the option of completing an additional fourth block of practice, which was also at 'game speed'. Participants were told that this optional block of practice was there to help them "*better prepare for the task*", but that it was not required.

Participants then completed three test blocks of the numerical discrimination task. Trials during the task followed the same basic set-up as during the practice trials, with fixation (250ms) followed by a dot image at 'game speed' (200ms), a response window of 1000ms and then performance feedback (850ms; Figure 1). At the end of the three test blocks participants were asked to estimate how well they performed on the task.

Following the ratings, participants were informed of an optional fourth block of the task. They were told that this block was not required, but that "*past research shows that practice helps on this task*". Before deciding whether to complete this optional block, participants were asked to rate how confident they were that they could improve their performance in the fourth block. Then participants chose whether to complete the optional fourth block of the task. After finishing the numerical discrimination task participants completed the family background demographic questionnaire (FBQ) and the abbreviated math anxiety scale (AMAS). Participants were debriefed at the end of the visit.

2.2.3 Measures

2.2.3.1 Numerical discrimination task. Participants were asked to discriminate between two overlapping sets of dots to determine whether there were a larger number of blue or orange dots on the screen. The two colors used for the different dot sets were luminance matched to be within 3 lux of each other. Half of the trials had more orange dots and half had more blue dots. All dot images fit into one of five different dot ratios, which included the ratios 1:2, 3:4, 5:6, 7:8 and 10:11. A total of 168 dot images were used from each of the five ratios, with 2 images from each ratio appearing in each practice block and 40 images from each ratio appearing in each task block. The total number of dots in each image ranged from 10 to 30. Half of the images were “dot-size controlled,” meaning that the average size of the dots in both sets were the same. In these trials the set with more dots covered more area on the screen. The other half of the trials were “area controlled,” meaning that the total area covered by each set of dots was the same. In these trials the larger set had smaller sized dots on average. A total of 840 images were used in this study, with 40 images appearing in the 4 practice blocks and 800 images appearing in the 4 task blocks. Dot images were presented in a pseudo-random order within each block. All participants viewed 630 images, with 36 participants completing the optional extra block of practice before starting the task (10 trials) and 57 participants completing an optional task block towards the end of the visit (200 trials). Dot images were created using Panamath version 1.22 (Halberda, Mazzocco, & Feigenson, 2008). The paradigm was run, and images were randomized using E-Studio 2.0 on HP Compaq 6910p laptop computers (see Figure 1).

2.2.3.2 Confidence measure. After consent was obtained and before the start of the task, participants were asked to predict how well they would do on the upcoming task using a 5-point likert-type scale ranging from '*very well*' (1) to '*very badly*' (5). They were then asked to rate their confidence in this prediction on a 5-point scale, where 1 was equivalent to '*very confident'* and 5 equaled '*very unsure*'. Participants were also asked to briefly explain their answers in an open response space.

After completing the first three blocks of the numerical discrimination task participants were asked to rate how well they thought they performed on a scale of 1 ('*very well*') to 5 ('*very badly*'), and how confident they were in that estimation from 1 ('*very confident*') to 5 ('*very unsure*'). Participants were then asked to explain the reasoning behind their ratings in a brief open-response section.

Finally, before deciding whether to complete the fourth block of the numerical discrimination task, participants were asked to rate how confident they were that they could better their previous performance on the task. The scores ranged from 1 ('*very confiden*t') to 5 ('*very unsure*'). Participants were also asked to briefly explain their reasoning. Scores were reversed during analysis for ease of interpretation, such that 5 was *'very confident/well'* and 1 was *'very unsure/badly'*.

2.2.4 Questionnaires

2.2.4.1 Depression, anxiety and stress scale short form (DASS-21). The DASS-21 is a valid and reliable measure of general depression, anxiety and stress (Henry & Crawford, 2005). The 14 questions from the DASS-21 pertaining to stress and anxiety were included on the pre-screen survey administered through the UMass SONA system prior to study participation. Participants used a 4-point likert-type scale to indicate how much each statement applied to them over the last week, ranging from '*did not apply to me at all' (0)* to '*applied to me very much, or most of the time*' (3). Scores for the relevant items were summed to provide a measure of participants' general stress and anxiety levels.

2.2.4.2 Abbreviated math anxiety scale (AMAS). The AMAS is a measure of math anxiety, which has shown good reliability and convergent/divergent validity (Hopko, Mahadevan, Bare & Hunt, 2003). The questionnaire consists of 9 math-related situations (i.e. "*Thinking about an upcoming math test 1 day before*") and participants indicate how anxious each situation would make them using a 5-point likert-type scale, ranging from '*low anxiety*' (1) to '*high anxiety*' (5). Responses are then summed to create a math anxiety score. The AMAS was administered twice over the course of this study, once during the SONA pre-screen and once after participants completed the numerical discrimination task in the laboratory.

2.2.4.3 Math, science and logic scale (MSLS). The MSLS consisted of 8 statements adapted from Brown & Josephs (2000), and Ben-Zeev, Dennehy, Sackman, Olides, & Berger (2011), and was designed to assess participant's identification with math and analytical reasoning. Participants indicated how strongly they agreed with each statement on a 9-point likert-type scale ranging from '*strongly disagree*' (1) to '*strongly agree*' (9). The MSLC was included on the SONA pre- screen survey. Answers were reverse-scored where appropriate (i.e. "Math abilities are not important to my success in school"), and summed to create a math identification score for each participant.

2.2.4.4 Family Background Questionnaire (FBQ). Participants completed a demographic questionnaire, after the numerical discrimination task, about their family background and their personal academic history. Questions covered topics such as the participants' ideal grade, average math and science grades and perceived effort in school, along with caregiver level of education, employment and basic information about family composition.

2.2.5 Statistical Approach. Descriptive statistics were calculated (See Table 1). To assess differences in task motivation separate chi-square measures were run. First the data were split by gender and two chi-square analysis were run (one within the females and one within the males) to assess the impact of condition on the decision to complete the fourth optional task round. Data were then split by condition and two chi-square analyses were run (one within the ST condition and one within the NT condition) to explore the impact of gender on the decision to complete the fourth round. Logistic regressions were used to determine which factors contributed to participant's decisions to complete the fourth task round. Separate analyses were run for male and female participants. Task motivation was entered as the dependent variable (coded as 1 for "yes to 4^{th} round" and 0 for "no to 4^{th} round). Task condition was entered as a covariate (coded as 1 for ST and 0 for NT). In one analysis math identification was entered as an independent variable, along with the interaction between math identification and condition, and percent correct entered as a control. In a second analysis, overall percent correct was entered as an independent variable, along with the interaction factor, and math identification score entered as a control.

For measures of task confidence univariate ANOVAs were used to determine any group differences. Multiple regressions were run to explore the impact of prior math identification on task confidence. Overall regressions were run with math identification scores and gender (coded as 0 for male and 1 for female), and condition (coded as 0 for NT and 1 for ST), along with the interaction variables were entered as independent variables, and overall task confidence was entered as the dependent variable. For follow-up analysis, this relation between math identification and confidence was also explored across conditions and within gender, with separate regression analyses run for male and female participants, and across genders, within conditions, with separate regressions run for NT and ST groups.

Multiple regressions were run to better understand the role of prior math identification on task performance, particularly during difficult trials. Overall regressions were run with math identification scores and gender (coded as 0 for male and 1 for female), and condition (coded as 0 for NT and 1 for ST), along with the interaction variables were entered as independent variables. In one analysis percent correct during difficult trials (ratios 7:8 and 10:11) was entered as the dependent variable. In a separate analysis percent correct during easy trials (ratios 1:2 and 3:4) was entered as the dependent variable. Follow-up regressions were run separately for the ST and NT conditions.

2.3 Results

2.3.1 Descriptive statistics. Participants were randomly assigned to either the ST group ("math intelligence") or the NT group ("creative ability"). Of the 62 female participants used in the analysis, 30 were assigned to the ST group and 32 were assigned to the NT group. Groups did not differ in age, $t(60)= .238$ $p= 0.813$, general anxiety, $t(59)= -.097$

p= 0.923, general stress, $t(59)$= -.588 p= 0.559, math anxiety level, $t(59)$= -1.629 p= 0.109, or math identification scores $t(50)$= -.678 p= 0.501. Of the 60 male participants, 31 were assigned to the ST group and 29 were assigned to the NT group. Similarly, these groups did not differ in age, $t(57)$= .197 p= 0.844, general anxiety levels, $t(57)$= .039 p= 0.969, general stress levels, $t(57)$= .872 p= 0.387, math anxiety level, $t(57)$= .180 p= 0.858, or math identification scores $t(57)$= -.807 p= 0.423 (See Table 1).

2.3.2 Task Motivation/Engagement. Chi-square analysis revealed that female participants in the ST group were much less likely to complete the optional fourth block of the numerical discrimination task compared to their peers in the NT group, $X^2(1)$ = 9.976, p = 0.002, φ=0.401. There was no significant group difference for the male participants (p=.809). Female participants were also much less likely to complete the optional fourth block compared to males in the ST condition, $X^2(1)$ = 7.878, p = 0.005, φ=-0.359. There was no gender difference among participants in the NT condition (p=.548; see Table 2).

Results from the logistic regressions analysis revealed that women in the ST condition, but not the NT condition, were more likely to complete the fourth round if they reported higher initial math identification, b=.726, SE = .343, p = .034 exp(B) = 2.066 (see Figure 2). Alternatively, women were more likely to complete the fourth block in the NT condition, but not in the ST condition, if they earned a higher percent correct during the first 3 blocks of the task, b=-.301, SE = .145, p = .038 exp(B) = .740 (see Figure 3). Neither percent correct nor math identification significantly impacted men's likelihood of completing the fourth round in either condition.

2.3.3 Task confidence. ANOVAs revealed a significant gender by condition interaction for overall task confidence, $F(1, 118) = 5.205$ $p = .024$ $\eta p^2 = .042$. Follow-up analysis revealed a gender difference in confidence ratings in the ST condition, such that men rated themselves as significantly more confident ($M = 3.5$ $SD = .56$) than women ($M = 3.0$ $SD = .59$), $F(1,59) = 8.338$ $p = .005$ $\eta p^2 = 124$. In contrast, there was no gender difference in confidence ratings for the NT condition ($p = .763$).

An overall condition by gender by math ID interaction was found for the regression analysis predicting overall task confidence, $b = .098$, $SE = .035$, $\beta = 4.223$, $p = .006$. Regression results indicate a significant interaction effect between gender and math identification in the ST condition ($b = .056$, $SE = .024$, $\beta = 2.699$, $p = .024$), but not the NT condition ($p = 109$). Simple slopes for the relation between math identification and confidence were calculated for males versus females in the ST group. Female participants showed a positive relation between math identification and confidence ($b = .057$, $SE = .021$, $\beta = .454$, $p = .012$). In contrast, male participants showed no significant relation between math identification and task confidence ($p = .971$; see Figure 4).

A significant condition by math identification interaction effect was found for female participants ($b = .086$, $SE = .029$, $\beta = 4.525$, $p = .005$), but not male participants ($p = .455$). Simple slope analysis revealed a significant positive relation between math identification and confidence for female participants in the ST condition (as seen above), but not in the NT condition ($p = .170$).

2.3.4 Task performance. No significant differences were found for task performance between groups. There was also no overall condition by gender by math ID interaction found for the regression analysis predicting percent correct ($p = .418$). However, exploratory analysis revealed a relation between math identification and task performance which differed across genders. An interaction effect between gender and math identification was found in the ST condition for difficult trials (*b*= -.550, *SE*=.221, β = -3.094, *p*=.016), but not for easy trials (*p*=.166). No interaction effects were found for the NT condition. Simple slopes analysis showed that women in the ST condition had a negative relation between math identification and performance on difficult trials (*b*= -.502, *SE*=.175, β = -.476, *p*=.008), whereas men had no significant relation between performance and math identification (*p*=.660; see Figure 5).

2.4 Discussion

This study aimed to better understand subtle stereotype threat (ST) effects on task motivation, confidence and performance. Further, we wished to explore the role of math identification as a possible protective factor. Our findings supported and expanded upon prior stereotype threat literature.

As hypothesized, women who were told that the task was a measure of “math intelligence” (ST condition) demonstrated lower task motivation compared to men in the same condition and women who were told the task measured “creative ability” (NT condition). Further, women in the ST condition reported lower confidence compared to men in the ST condition. As the task remained the same across conditions and genders, the lower task confidence and the choice to forgo the fourth block were not due to the task itself, but rather to how the context of the task was framed. These findings are consistent with prior literature, suggesting that women disengage from a threatening task as a self-protective measure when under ST (Beasley & Fischer, 2012; Steele, 1997); the results also lend support to the notion that at least part of women’s lower motivation to enter STEM-related fields is driven by subtle ST effects rather than by an inherent disinterest in the work.

Importantly, our study found that women under ST who reported high math identification prior to the study were more likely to continue the task and report higher task confidence than those with lower math identification. These findings support previous literature demonstrating that domain identification is positively related to intrinsic motivation (Walker, Greene & Mansell, 2006), and suggest that math identification may help protect women's drive against ST effects. Interestingly, women's task motivation in the NT group, but not the ST group, was predicted by their task performance, with higher performing women more likely to complete an extra round. This finding suggests that when women are not under the stressor of ST they can effectively track their performance and decide whether to continue a task accordingly. However, women under ST may lose this ability, as ST is thought to bias attention towards negative feedback (Forbes & Leitner, 2014) and impair working memory (Schmader & Johns, 2003). Thus, women under ST may rely on prior protective factors, such as domain identification, to motivate task engagement, rather than immediate performance.

Our study found no difference in overall accuracy between the groups. The lack of group differences in performance may be due to the use of the numerical discrimination task, which has not previously been used in studies of ST. It seems likely that the relatively novel aspects of the numerical discrimination task reduced any advantage that male participants may have had in a more traditional math task. For instance, male students tend to have had more traditional math experience and explicit math interest during late high school and college, compared to their female peers (Ceci & Williams, 2010; Sadler et al., 2012). Therefore, gender differences in traditional math tasks may be impacted by factors outside of the study's stereotype manipulation, (e.g. practice effects among the male participants), and these differences may have been avoided with the use of the numerical discrimination task which is novel in ST studies. Further, the numerical discrimination task requires less use of executive functioning (i.e. working memory, selective attention) than typical math tasks as it relies on numerical approximations rather than calculations (Cragg & Gilmore, 2014; Halberda, Mazzocco, & Feigenson, 2008). As ST is thought to negatively impact performance at least partially by exhausting executive functioning (Rydell et al., 2009; Schmader et al., 2008), overall performance under ST may not be hindered to the same extent on tasks that do not require these skills.

Although there was not an overall group difference in performance, there was a negative relation between math identification and task performance on difficult trials for women in the ST group. This finding supports previous literature demonstrating that women under ST perform worse, particularly on difficult task items, if they report higher domain identity (Keller, 2007). This finding, in combination with the motivation and confidence results discussed above, suggest that high domain identity is simultaneously a vulnerability and a potential protective factor for individuals under ST. Interestingly, ST only strongly impacts performance for individuals who wish to disprove the negative stereotypes about their in-group (Aronson et al., 1998; Dasgupta, 2011). Thus, individuals with high domain identification are more susceptible to ST-induced stress and subsequent performance deficits. However, high domain identification also increases motivation within the stereotyped task, regardless of current task performance. Over time, the increased vulnerability to ST-induced stress may overpower the performance motivation effects of high domain identity. This relation would explain Osborne and Walker's finding that African American students with higher academic identity are both more likely to have higher GPAs and to subsequently withdraw from school (2006). This combination of effects (increased motivation to succeed within ST domains and increased susceptibility to ST-induced stress) indicates that future ST intervention work should focus on increasing domain identification at a young age, before the build-up of stressors can push these individuals towards domain de-identification. Furthermore, interventions seeking to decrease the stress effects of ST should focus on individuals with pre-existing high domain identification, as they are the ones most likely to benefit from these efforts.

The current study has a few limitations. For one, this study focuses on a population of students from the Psychological and Brain Sciences department within the College of Natural Sciences. Future studies should explore whether these ST effects remain stable across different types of STEM disciplines, including fields such as chemistry and engineering. Further, due to the nature of the local population, the sample was relatively homogenous in race, with the largest subsection reporting as Caucasian/white (approximately 64%). Future research is needed to explore the intersection of ST effects in more racially diverse populations.

Despite these potential limitations, this study advances our understanding of ST effects on women's motivation and confidence within STEM-related tasks and the potential protective influence of domain identification. These findings support the notion that subtle ST negatively impacts women's choice to continue tasks linked to threatening fields, separate from any impact of the tasks themselves. These findings demonstrate that what may otherwise be viewed as women's "free choice" to avoid STEM fields is impacted by subtle societal expectations. Further, our results show that high pre-existing domain identification may help protect women against these ST effects on task motivation and confidence. Future efforts to lessen the gender gap in STEM fields should focus on increasing domain identification at an early age, and reducing long-term ST stress effects, thus increasing women's overall motivation to succeed within these stereotyped domains.

Table 1. Means and standard errors for age, anxiety scores, stress scores, math anxiety level and math identification scores in Study 1

	Female		Male	
	Stereotype Threat	Non-Threat	Stereotype Threat	Non-Threat
Final Sample Size	30	32	31	29
Age: Mean *(SE)*	19.73*(0.26)*	19.81*(0.21)*	19.55*(0.20)*	19.61*(0.22)*
Anxiety: Mean *(SE)*	9.07*(1.47)*	8.84*(1.81)*	7.42*(1.59)*	7.50*(1.25)*
Stress: Mean *(SE)*	12.47*(1.47)*	11.10*(1.80)*	1.48*(1.77)*	9.64*(1.72)*
Math Anxiety Level: Mean *(SE)*	2.17*(0.14)*	1.87*(0.12)*	1.58*(0.10)*	1.61*(0.11)*
Math ID: Mean *(SE)*	57.53*(0.86)*	56.75*(0.78)*	52.39*(1.72)*	50.43*(1.71)*

Table 2. Number of participants who chose to complete the optional fourth task round per group in Study 1.

	Female		Male	
	Stereotype Threat	Non-Threat	Stereotype Threat	Non-Threat
Yes to 4th Round	6	19	17	15
No to 4th Round	24	13	14	14

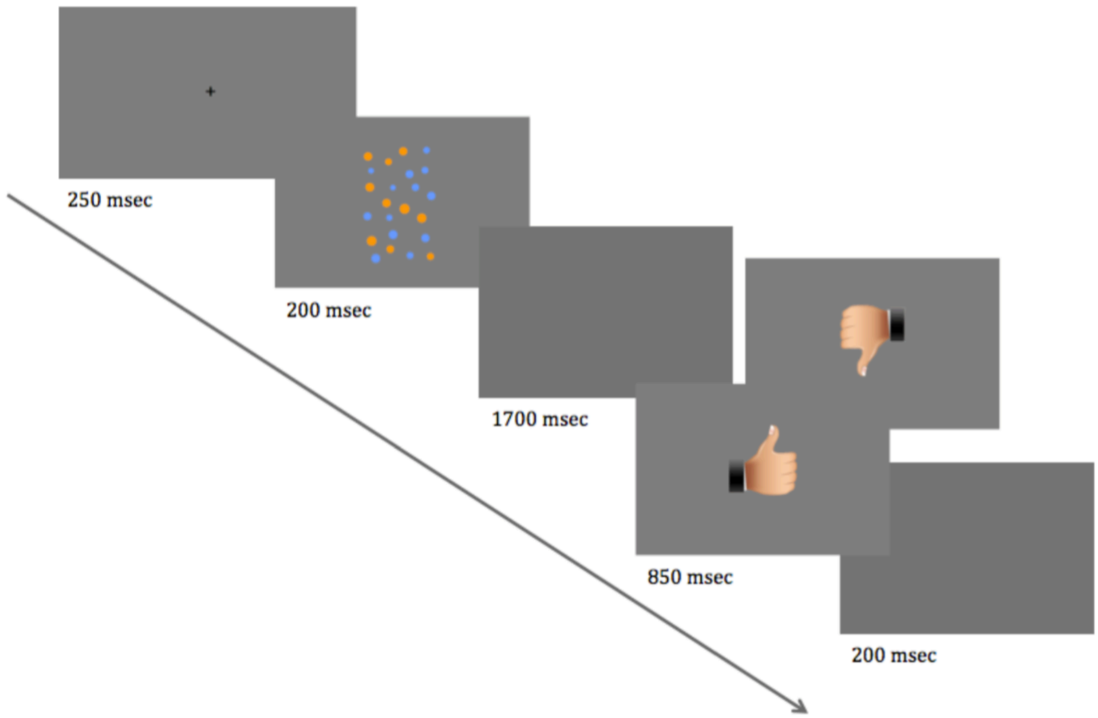


Figure 1. Numerical discrimination task sequence.

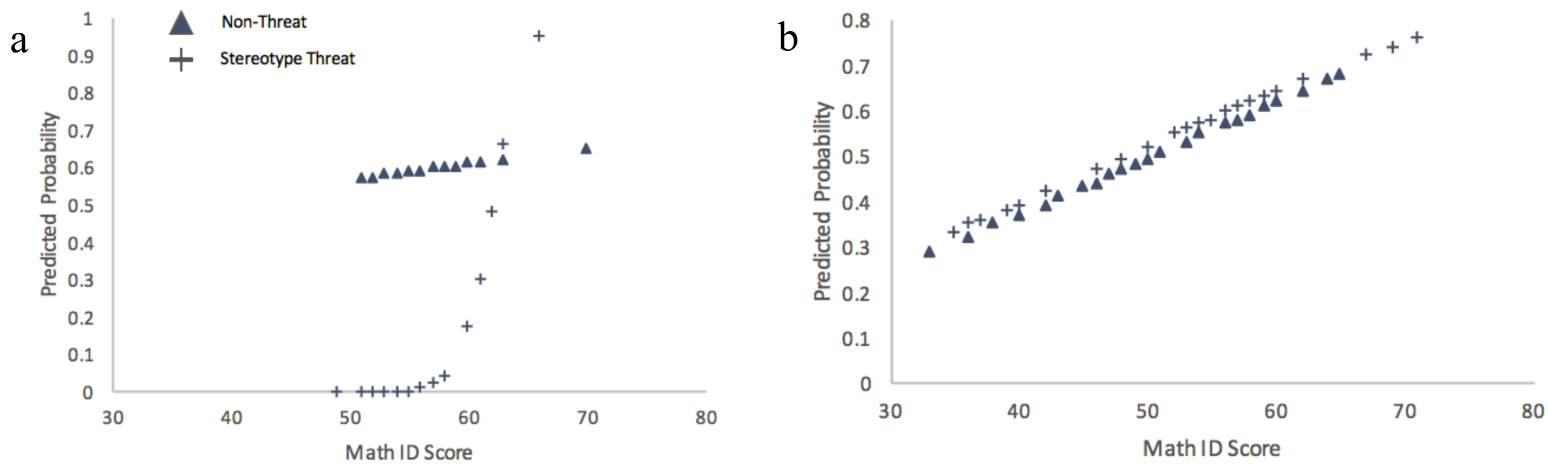
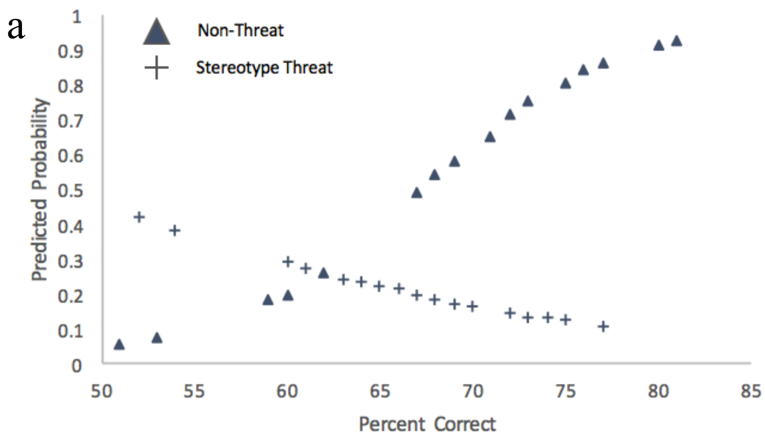


Figure 2. Logistic regression of math identification and condition. The patterns show predicted probability of choosing to complete the 4th round for a) female participants, and b) male participants.



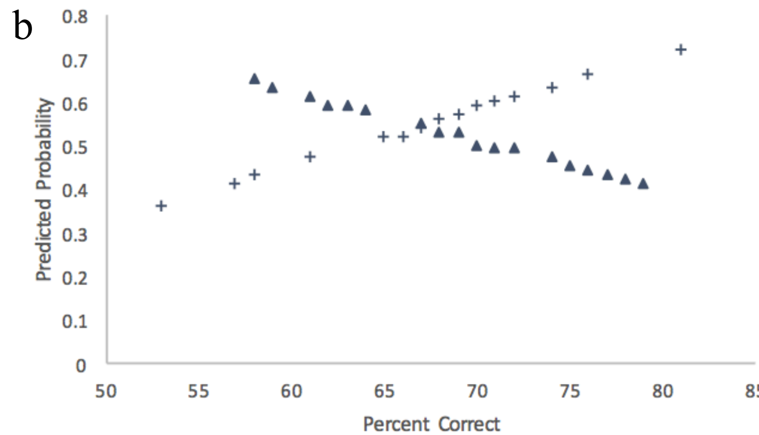


Figure 3. Logistic regression of percent correct and condition. The patterns show predicted probability of choosing to complete the 4th round for a) female participants, and b) male participants.

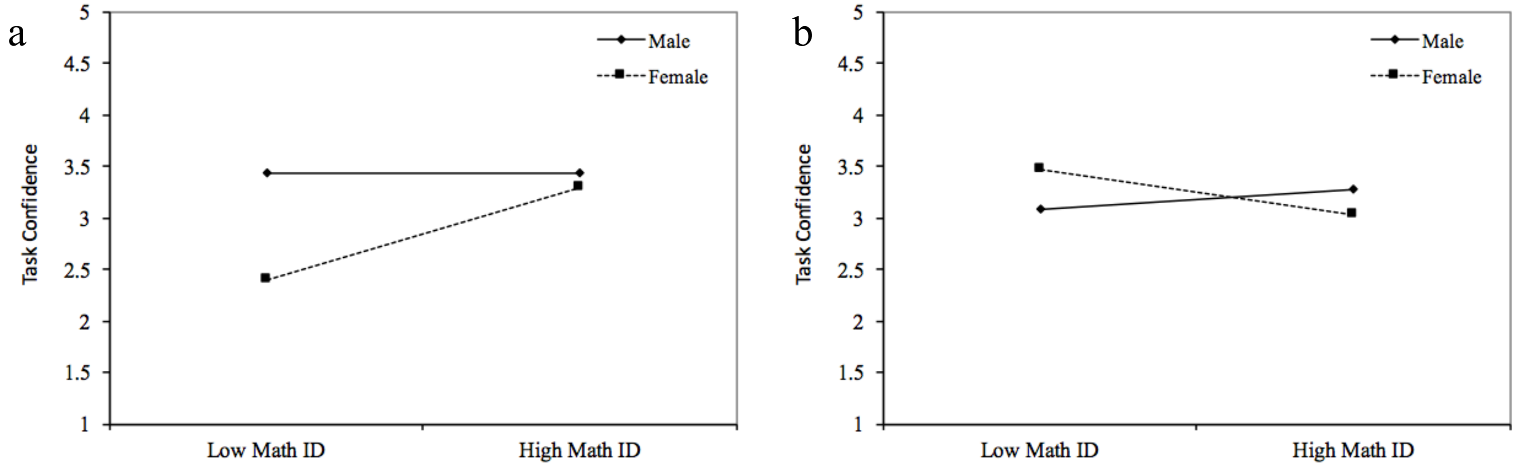


Figure 4. Predicted relation between task confidence, gender and math identification for a) participants in the ST condition and b) participants in the NT condition.

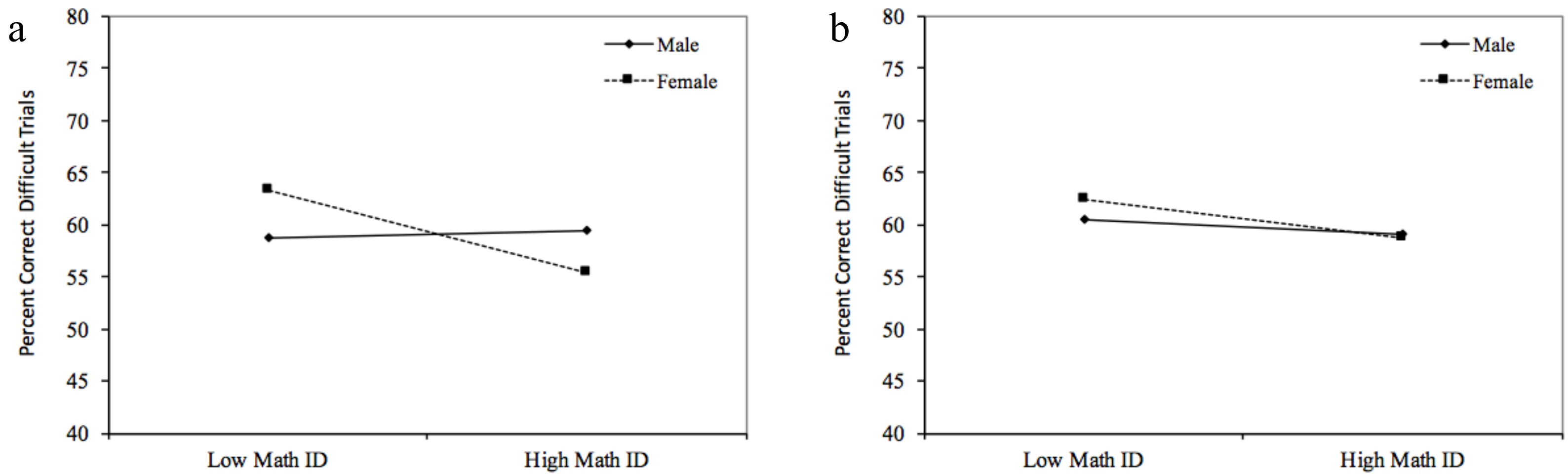


Figure 5. Predicted relation between percent correct on difficult trials, gender and math identification for a) participants in the ST condition and b) participants in the NT condition.

CHAPTER 3

STUDY 2: NEURAL MARKERS OF PERFORMANCE MONITORING UNDER ST

3.1 Aims and Hypothesis

Study 2 aimed to explore the impact of ST on young women's performance monitoring. We hypothesized that women in the ST group would demonstrate enhanced error detection (larger error-related negativity/ERN) during easy trials, when internal performance monitoring is plausible, and enhanced feedback processing (larger FRN) during difficult trials, when internal performance monitoring is more difficult. We also hypothesized that participants would show less neural processing of errors (attenuated P300 and error positivity/Pe) when told the task measures math ability (ST group) compared to the NT group. This pattern of lower neural processing of errors would account for increased error saliency during ST conditions, combined with a difficulty to effectively use error information to improve task performance. Further, we predicted that math identity would relate to larger ERN amplitudes during easy trials and larger FRN amplitudes to negative feedback during difficult trials in the ST group but not in the NT group; we predict this pattern based on the idea that math ID may enhance math-based ST effects, undermining immediate task performance and generating less adaptive performance-monitoring.

3.2 Methods

3.2.1 Participants. To participate, students needed to be at least 18 years of age, and they needed to have completed the pre-screen survey via the UMass Amherst participant recruitment SONA System. Students could not participate if they had a diagnosed learning or attention disability, or if they were colorblind. Students were compensated for their time with extra credit points in participating psychology courses.

A total of fifty-eight female college students were recruited for Study 2. Of these participants ten were excluded due to equipment failure, one was excluded because of experimenter error, three correctly guessed the study manipulation, and four were excluded from analysis because they were at least 3 standard deviations away from the group on key variables. Overall, excluded participants did not differ from included participants on general anxiety, $t(55) = -.055$ $p=.956$, general stress, $t(55) = -.422$ $p=.675$, math identification score, $t(54) = -.753$ $p=.455$, or distribution amongst the two conditions, $X^2(1) = .718$, $p = 0.397$, $\varphi=.113$. Excluded participants did differ in age from included participants, $t(56) = 3.038$ $p=.004$, with included participants being older on average ($M= 19.33$, $SE= .166$) than excluded participants ($M= 18.50$, $SE= .167$). However, age did not differ between the various reasons for exclusion, $F(3,14) = .251$ $p=.860$. Participants in Study 2 also did not differ from female participants in Study 1 in age, $t(100) = 1.639$ $p=.104$, general stress, $t(98) = 1.192$ $p=.236$, general anxiety, $t(98) = .787$ $p=.433$, or math identification, $t(98) = .607$ $p=.545$.

3.2.2 Procedure. Participants were randomly assigned to either the ST group or the NT group upon arrival. Participants were brought into the lab one at a time, where a male research assistant greeted them. Following consent, the participant was fitted with a cap for EEG/ERP processing. The rest of the procedure remained as described in Study 1.

3.2.3 Psychophysiological recording and data reduction. EEG was recorded using Ag-AgCl electrodes in a 64-channel Lycra Electro-Cap setup in accordance with the International 10-20 System. Eye movements were regressed from the data. Mastoid electrodes served as reference and impedances were kept < 20kΩ. Data were filtered (0.01-100Hz), amplified and digitized (1000Hz), and then filtered again during processing with a 30Hz low-pass. The EEG was baseline corrected, and trials containing artifacts (i.e., epochs exceeding an EEG voltage threshold of +/-150μV) were removed.

ERPs were constructed separately for components that were time-locked to response onset (i.e., the ERN, CRN, Pe $_{(incorrect)}$, and Pe$_{(correct)}$) and those time-locked to feedback onset (i.e., FRN$_{(negative)}$, FRN$_{(positive)}$, P300$_{(negative)}$, and P300$_{(positive)}$). All components were baseline corrected using a window of 200ms prior to response onset or feedback onset, respectively.

Response-locked ERPs*:* In accordance with inspection of the grand means and based on previous literature (e.g., Forbes et al., 2008; Luu, Tucker & Makeig, 2004), the peak amplitude for the ERN and CRN components was scored as the negative most peak between 0-100ms after incorrect and correct responses, respectively, at the frontral-central region (sites FC1, FCz and FC2). The Pe component was maximal at the parietal region (sites P1, PZ and P2; also see Orr & Carrasco, 2011; Forbes et al., 2008; Falkenstein, Hohnsbein, Hoormann & Blanke, 1991) and was scored as the most positive peak between 50-250ms for incorrect (Pe$_{(incorrect)}$) and correct (Pe$_{(correct)}$) responses. For each Pe component, a peak to peak score was calculated by subtracting the preceding negative component (i.e., Pe$_{(incorrect)}$ peak-to-peak score = Pe$_{(incorrect)}$ minus ERN, and the Pe$_{(correct)}$ peak-to-peak score = Pe$_{(correct)}$ minus CRN).

Response-locked ERP difference scores were also created in accordance with methods used in previous literature (Xiao et al., 2011; Talmi, Fuentemilla, Litvak, Duval

& Dolan, 2012). Difference scores controlled for reactivity during correct trials during incorrect trial analysis (i.e., ERN minus CRN; $Pe_{(incorrect)}$ minus $Pe_{(correct)}$).

Feedback-locked ERPs*:* The FRN peak amplitude was scored as the most negative peak within the 200-375ms window following feedback onset, with the $FRN_{(negative)}$ following negative feedback and the $FRN_{(positive)}$ following positive feedback. The final FRN peak values used in analysis were computed as a peak-to-peak score between the preceding positive P200 (scored as the most positive peak in a window of 50-250ms following feedback onset) and the FRN ($FRN_{(negative)}$ peak-to-peak = $FRN_{(negative)}$ peak – $P200_{(negative)}$ peak; $FRN_{(positive)}$ peak-to-peak = $FRN_{(positive)}$ peak – $P200_{(positive)}$ peak. Based on inspection of the grand mean waveforms as well as previous studies (e.g., Mangels et al., 2012; Masser, Rossi, Schutter & Kenemans, 2012). The FRN analysis focused on the frontal-central region (sites FC1, FCz and FC2).

The P300 was scored as the most positive going peak from 250-500ms post feedback onset, with the $P300_{(negative)}$ following negative feedback and the $P300_{(positive)}$ following positive feedback. The P300 values used in analysis were the peak-to-peak scores between the respective preceding FRN peak and the P300 peak (i.e. the peak to peak $P300_{(negative)}$ score = $P300_{(negative)}$ minus $FRN_{(negative)}$ and the peak to peak $P300_{(positive)}$ score = $P300_{(positive)}$ minus $FRN_{(positive)}$). The P300 analysis focused on the parietal region (sites P1, Pz and P2) in accordance with grand mean waveforms and prior literature (i.e., Xu, Shen, Chen, Ma, Sun & Pan, 2011; Schaefer, Buratto, Goto & Brotherhood, 2016).

Feedback-locked ERP difference scores were also calculated in accordance with methods used in previous literature (Xiao et al., 2011; Talmi, Fuentemilla, Litvak, Duval & Dolan, 2012). Difference scores allowed for analysis of ERP reactivity to negative feedback, controlling for reactivity to positive feedback (i.e. FRN difference score = $FRN_{(negative)}$ minus $FRN_{(positive)}$; P300 difference score = $P300_{(negative)}$ minus $P300_{(positive)}$).

3.2.4 Measures

3.2.4.1 Numerical discrimination task. The overall task remained the same as described in Study 1. However, to lessen the time that the task took, the most difficult dot ratio (10:11) was removed from testing. Therefore, all dot images fit into one of four different dot ratios, including the ratios 1:2, 3:4, 5:6, and 7:8. A total of 168 dot images were used from each of the four ratios, with 2 images from each ratio appearing in each practice block and 40 images from each ratio appearing in each task block. All controls remained as described in Study 1.

3.2.4.2 Confidence measure. The confidence measure remained as described in Study 1. In sum, participants were asked to predict how well they would do on the upcoming task after ST manipulation but before they saw the task itself. They were then asked to rate their confidence in this prediction and to briefly explain their answers. After completing the first three blocks of the numerical discrimination task participants were asked to rate how well they thought they performed, how confident they were in that estimation, and to explain the reasoning behind their ratings in a brief open-response section. Before deciding whether to complete the fourth block of the numerical discrimination task, participants were asked to rate how confident they were that they could better their previous performance on the task, and to briefly explain their reasoning. Finally, a subset of participants were asked to rate the task on a 5-point Likert scale, with 1 meaning the task felt more like a game and 5 meaning the task felt more like a test.

3.2.4.3 Questionnaires. The questionnaires remained as described in Study 1. As before, participants completed the DASS-21, the MSLS, and the first AMAS during the SONA prescreen, and the second AMAS and the FBQ after task performance.

3.2.5 Statistical approach. Descriptive statistics examining participant attitudes and age were assessed using independent t-tests. To determine whether there were any group differences in the number of epochs for response-locked and feedback-locked waveforms a series of Univariate ANCOVAs, controlling for age, math anxiety, time of day (as time-of-day has been shown to impact attention; Matchock, R.L., and Mordkoff, J.T., 2008), and average reaction time (as longer reaction times may have begun to exceed the response time and as such coded as errors of omission), were conducted.

A series of Univariate ANCOVAs, controlling for age, math anxiety and percent correct where appropriate, were used to explore group differences in behavioral measures. Regression analysis, controlling for age and math anxiety, was used to examine potential group differences in the relation between attitudes (i.e., general anxiety, general stress) and task performance. Chi-square analysis was used to find any group differences in task motivation as reflected by the decision to complete the fourth round. Logistic regressions were used to determine which factors contributed to decisions to complete the fourth round. For the logistic regression, confidence at the start of the task and percent correct were entered as dependent variables and the fourth round decision was entered as the independent variable (0 for "no fourth round"; 1 for "yes fourth round").

ERP amplitudes were first assessed via a series of 2x2x2 repeated measures analysis of covariance (ANCOVAs), controlling for age, math anxiety and percent correct. Level (easy vs. hard) and trial type (incorrect vs. correct) were entered as within-subject factors. Condition (ST vs. NT) was entered as the between-subjects factor. Separate analyzes were performed for each set of ERP components (i.e., ERN and CRN, $Pe_{(incorrect)}$ and $Pe_{(correct)}$, $FRN_{(positive)}$ and $FRN_{(negative)}$, $P300_{(positive)}$ and $P300_{(negative)}$).

Next ERP difference scores were analyzed with a series of 2x2 ANCOVAs were performed, controlling for age, math anxiety and percent correct. Level (easy vs. hard) was entered as the within-subjects factor and condition (ST vs. NT was entered as the between-subjects factor. Separate analyzes were performed for each ERP component difference score. To examine significant patterns that emerged in the repeated measures ANCOVAs, follow-up univariate ANCOVAs and t-tests were used.

Multiple regressions were run to explore the impact of attitudes (math identification, math anxiety, general anxiety, general stress, task confidence) on response-locked ERPs, controlling for age in between-group comparisons, and controlling for percent correct where appropriate. Regressions were also run to explore the relation between ERP reactivity and task performance (i.e., percent correct, errors of omission, post-error slowing), controlling for age and math anxiety. In one set of regressions, ERPs to correct and incorrect trials were examined separately (i.e. $FRN_{(positive)}$ was entered separately from $FRN_{(negative)}$). In a second set of regressions, ERP difference scores between correct and incorrect trials were explored (i.e. $FRN_{(negative)}$ minus $FRN_{(positive)}$).

As previous literature suggested a possible inverse relation between ERN and Pe amplitudes under stress (Moser, Moran & Jendrusina, 2012; Hajcak, McDonald & Simons, 2004), multiple regressions were used to explore the relation between ERN amplitude and $Pe_{(incorrect)}$ reactivity, and between CRN and $Pe_{(correct)}$ reactivity. A negative relation between ERN and FRN amplitudes was previously demonstrated (Heldmann, Russeler and Munte, 2008; Stahl, 2010). To explore the possibility of this relation in the current study, regressions were used to examine the relation between ERN and $FRN_{(negative)}$ reactivity, and CRN reactivity and $FRN_{(positive)}$ reactivity. In all regressions age, math anxiety and percent correct were controlled for.

3.3 Results

3.3.1 Descriptive statistics. Participants were randomly assigned to either the ST group ("math intelligence", n= 22) or the NT group ("creative ability", n=18). Groups did not differ in general anxiety, $t(37)= -1.246$ $p= 0.220$, general stress, $t(37)= -1.541$ $p= 0.132$, or math identification scores $t(36)= -1.532$ $p= 0.134$ (see Table 3a), or the number of usable epochs for response-locked or feedback-locked components (see Table 3b). Groups did show a difference in age, $t(38)= 2.001$ $p= 0.05$, and a trending difference in math anxiety level, $t(38)= -1.816$ $p= 0.077$, so both measures were controlled for in the following analysis.

3.3.2 Task Behavior. No group differences in percent correct (p=.693), post-error slowing (p=.833), or errors of omission (p=.677) were found. There was also no group difference in task confidence (p=.477).

However, regressions revealed a group difference in the relation between general anxiety and percent correct, b= -.747, SE=.244, β = -1.034, p=.005. Simple slopes analysis demonstrated a negative relation within the ST condition, b= -.442, SE=.147, β = -.665, p=.008, such that higher general anxiety predicted worse task performance. No relation between general anxiety and accuracy emerged for the NT condition (p=.907). Another group difference was found for the relation between general stress and post error slowing, b= -4.862, SE=1.535, β = -1.113, p=.003. Simple slopes analysis showed a significant negative relation within the ST condition, b= -4.863, SE=1.632, β = -1.037, p=.009, with higher stress scores predicting less post-error slowing. Stress scores were not related to post-error slowing in the NT group (p=.571; see Figure 6).

3.3.3 Task motivation/engagement. Chi-square analysis revealed no difference between the groups in the decision to complete the optional fourth block of the task (p=.257; see Table 4). Regression results indicate that women with lower initial confidence were more likely to complete the 4th round, regardless of condition, b=1.245, SE = .550, p = .024 *exp*(B) = 3.472 (see Figure 7).

3.3.4 ERN and CRN. A trend Level x Condition interaction (F(1, 33) = 3.783 p = .060) emerged. Follow-up analysis indicated no group difference in neural reactivity (collapsed across ERN and CRN components) at the easy (p= .188), or hard level (p= .886). However, paired t-tests within each group demonstrated a significant Level difference in neural reactivity in the ST group, t(20) = -2.735 p=.013, such that participants reacted more strongly (i.e., had larger neural responses) to the easy trials (M=-2.767 SE= .287) compared to the difficult trials, (M=-1.645 SE= .277). There was no within group difference for Level in the NT group (p=.976; see Figure 8). For the ERN minus CRN difference score, there were no Condition differences (p=.246) or Level x Condition differences (p=.422).

No relation emerged for either ERN or CRN amplitudes and performance measures (see Tables 5a and 5b, respectively), or between the ERN-CRN difference score and performance measures (see Table 5c).

No relation was found between either ERN or CRN reactivity and participant attitudes (see Tables 6a and Table 6b, respectively). The ERN-CRN difference score during difficult trials related to math identification differently by group, b= 5.255, SE=2.309, β = .564, p=.030. Simple slopes analysis revealed that in the ST group, higher math ID scores related to more attenuated neural reactivity during difficult trials, b= 4.356, SE=1.706, β = .554, p=.022. No relation was found within the NT group between math identification and the ERN-CRN difference score to difficult trials (p=.647). Likewise, no group difference was found for the relation between the ERN-CRN difference score to easy trials and math ID (p=386; see Table 6c).

A group difference was also found for the relation between the ERN-CRN difference score to difficult trials and general anxiety, b= 7.006, SE=2.552, β = .587, p=.010. Simple slopes demonstrated a positive relation in the ST group, with higher anxiety predicting more attenuated neural reactivity during difficult trials, b= 3.739, SE=1.454, β = .352, p=.021. No relation was found in the NT group (p=.291). There was also a significant group difference in the relation between the ERN-CRN difference score to difficult trials and general stress, b= 8.009, SE=2.891, β = .596, p=.009. Simple slopes analysis demonstrated a positive relation in the ST group, with higher stress predicting more attenuated neural reactivity during difficult trials, b= 5.77, SE=1.63, β = .484, p=.003 (see Figure 9). No other relations were found for the ERN-CRN difference score and participant attitudes (see Table 6c).

3.3.5 Pe. A 3-way interaction was found between Trial Type (Incorrect vs. Correct) x Level (Easy vs. Hard) x Condition (ST vs. NT), $F(1, 30) = 4.42$ $p = .044$ for the Pe difference score at the parietal region. Similarly, a 2-way Level x Condition interaction was found for the Pe difference score ($Pe_{(incorrect)}$ - $Pe_{(correct)}$ $Pe_{(incorrect)}$ - $Pe_{(correct)}$, $F(1, 30) = 4.42$ $p = .044$.

As follow-up for the 3-way interaction, paired t-tests within the ST group revealed a significant difference in $Pe_{(incorrect)}$ compared to $Pe_{(correct)}$ amplitudes during easy trials, $t(19) = 4.74$ $p < .001$. In contrast, paired t-tests within the NT group showed a significant difference in $Pe_{(incorrect)}$ compared to $Pe_{(correct)}$ amplitudes during difficult trials, $t(16) = 3.04$ $p = .008$.

A significant group difference was found for the relation between $Pe_{(incorrect)}$ amplitudes during easy trials and errors of omission, $b = 8.286$, $SE = 3.998$, $\beta = .968$, $p = .046$. Simple slopes revealed a significant positive relation within the ST group, $b = 8.604$, $SE = 3.145$, $\beta = .595$, $p = .014$, such that higher $Pe_{(incorrect)}$ during easy trials predicted more errors of omission. There was no relation between $Pe_{(incorrect)}$ amplitudes and errors of omission in the NT group ($p = .91$).

Similarly, there was a group difference in the relation between $Pe_{(correct)}$ during easy trials and errors of omission, $b = 9.445$, $SE = 3.663$, $\beta = .663$, $p = .015$. Simple slopes revealed a significant relation within the ST group between $Pe_{(correct)}$ amplitudes during easy trials and errors of omission, $b = 9.701$, $SE = 3.300$, $\beta = .603$, $p = .009$, with higher $Pe_{(correct)}$ amplitudes predicting more errors of omission. No relation was found between $Pe_{(correct)}$ and errors of omission in the NT group ($p = .723$).

A significant group difference was found for the relation between $Pe_{(incorrect)}$ amplitudes during difficult trials and percent correct, $b = -3.12$, $SE = 1.15$, $\beta = -1.04$, $p = .011$. Simple slopes revealed a significant negative relation within the ST group, $b = -1.72$, $SE = .57$, $\beta = -.52$, $p = .01$, with lower $Pe_{(incorrect)}$ amplitudes during difficult trials predicting higher percent correct. No significant relation was found for the NT group (*p*=.52). No other relations were found between $Pe_{(incorrect)}$ and performance measures (see Table 7a).

A group difference was found for the relation between $Pe_{(correct)}$ amplitudes to easy trials and percent correct, *b*= -1.34, *SE*=.55, β = -.41, *p*=.02. Simple slopes showed a significant relation within the ST group, *b*= -1.75, *SE*=.54, β = -.53, *p*=.01, with more $Pe_{(correct)}$ reactivity predicting lower percent accuracy. No relation was found for the NT group (*p*= .83). Similarly, a significant group difference was found for the relation between $Pe_{(correct)}$ amplitudes to easy trials and post-error slowing, *b*= -14.13, *SE*=4.89, β = -.62, *p*=.01. Simple slopes showed a significant relation within the ST group, *b*= -13.94, *SE*=4.87, β = -.54, *p*=.012, with less $Pe_{(correct)}$ reactivity predicting more post-error slowing. No such relation was found for the NT group (*p*=.897). Finally, there was a group difference in the relation between $Pe_{(correct)}$ amplitudes to easy trials and errors of omission, *b*= -9.63, *SE*=3.46, β = -.63, *p*=.01. Simple slopes showed a significant positive relation within the ST group, *b*= 9.56, *SE*=3.35, β = .61, *p*=.01, with heightened $Pe_{(correct)}$ amplitudes predicting more errors of omission. No group differences were found for the relation between $Pe_{(correct)}$ amplitudes during difficult trials and performance measures (see Table 7b). Regressions revealed no relations between $Pe_{(incorrect)}$ to $Pe_{(correct)}$ difference scores and performance measures (i.e., errors of omission, percent correct, post-error slowing; see Table 7c).

There was no relation between $Pe_{(incorrect)}$ amplitudes and participant attitudes (math identification, general anxiety, general stress, task confidence; see Table 8a). Similarly, there was no relation between $Pe_{(correct)}$ and participant attitudes (see Table 8b) or between

$Pe_{(incorrect)}$ minus $Pe_{(correct)}$ difference scores and participant attitudes (see Table 8c).

3.3.6 FRN. There were no group differences in $FRN_{(negative)}$ or $FRN_{(positive)}$ peak-to-peak scores at the frontal central region. For FRN difference scores there were no Condition differences (p=.357) or Level x Condition differences (p= .736) in .

There was no relation between $FRN_{peak\text{-}to\text{-}peak}$ amplitudes and task performance (percent correct, errors of omission or post-error slowing; see Tales 9a and 9b respectively). A significant relation was revealed between the FRN difference score ($FRN_{(negative)}$ minus $FRN_{(positive)}$) reactivity during easy trials and overall percent correct, b= -1.675, SE=.544, β = -.501, p=.004, such that higher FRN amplitude to easy errors were related to greater overall task accuracy. There was also an overall relation between FRN difference scores during easy trials and total errors of omission, b= -9.09, SE=3.22, β = -.44, p=.008, with higher FRN difference score amplitudes to errors predicting more errors of omission (see Table 9c).

There was no relation between $FRN_{(negative)}$ peak-to-peak amplitude and prior attitudes (Math ID, general anxiety, general stress scores, or task confidence; see Table 10a). There was a positive relation across groups between $FRN_{(positive)}$ peak-to-peak amplitude during easy trials and overall task confidence such that greater $FRN_{(positive)}$ reactivity predicted higher task confidence, b= .095, SE=.045, β = .350, p=.042. No other relations were found between $FRN_{(positive)}$ reactivity and participant attitudes (see Table 10b). Regressions, controlling for age and percent correct revealed no relation between FRN difference scores and participant attitudes; see Table 10c).

3.3.7 P300. There were no group differences for the P300 peak-to-peak values at the parietal region. There were also no Condition differences (p=.593) or Level x Condition differences (p= .155) in P300 difference scores ($P300_{(negative)}$ minus $P300_{(positive)}$).

No overall relation was found between $P300_{(negative)}$ or $P300_{(positive)}$ peak-to-peak amplitudes and task performance (percent correct, errors of omission or post-error slowing; see Table 11a and Table 11b respectively). A significant relation was revealed between the P300 difference score to negative feedback on difficult trials and overall percent correct, b= 1.24, SE=.43, β = .43, p=.007. Specifically, higher P300 reactivity to difficult negative feedback predicted lower percent correct (see Table 11c).

Regressions revealed no overall relation between P300 peak-to-peak scores or P300 difference scores and participant attitudes (math identification, general anxiety, general stress, task confidence; see Tables 12a,12b and 12c).

3.3.8 Association between ERN, CRN and Pe. No relation was found between ERN and $Pe_{(incorrect)}$ amplitudes at easy (p=.75) or hard levels (p=.21). No relation was found between CRN and $Pe_{(correct)}$ amplitudes at easy (p=.77) or hard levels (p=.90). Similary, there was no significant difference between ERN difference scores and Pe difference scores at easy (p=.95) or hard levels (p=.23).

3.3.9 Associations between ERN, CRN and FRN. A group difference was discovered for the relation between the ERN difference score and the FRN difference score for easy errors, b= -1.174, SE=.450, β = -.938, p=.014. Specifically, participants in the NT group demonstrated a positive relation between these components, such that more neural reactivity to easy errors (ERN difference score) predicted more neural reactivity to negative feedback after easy errors (FRN difference score), b= 1.024, SE=.370, β = .566, p=.017. There was no significant relation between ERN difference scores and FRN difference scores in the ST group (p=.374; see Figure 11).

3.3.10 Mean score analysis. Overall ANCOVAs were run to determine whether major group findings within the peak and peak-to-peak scores remained the same when measured as mean score. The condition by level findings for the ERN and CRN peak scores did not hold with mean scores, $F(1, 33) = .521$ $p = .475$. Further, the condition by trial type by level interaction found for the Pe peak-to-peak scores did not hold when mean scores were used, $F(1, 30) = 2.287$ $p = .141$.

3.4 Discussion

This study aimed to understand the impact of ST on performance monitoring by using ERPs to measure temporally sensitive changes in cognitive processing that may not be apparent in more explicit measures. The analysis focused on a set of ERP components strongly linked with internal response-monitoring (i.e., ERN/CRN and Pe) and external feedback-monitoring (i.e., FRN and P300). Our results suggest that ST effects may impact internal response-monitoring processes more robustly than external feedback-monitoring processes.

Performance differences were explored between the two conditions to examine the ST manipulation. Although there were no overall group differences in performance measures (i.e., percent correct, errors of omission or post-error slowing), there were group differences in the relation between these measures and participant proneness to anxiety and stress. Specifically, women in the ST condition showed a negative relation between anxiety and accuracy rate (i.e. percent correct), and a negative relation between stress and post-error slowing, with no such relation apparent in the NT condition. These patterns are consistent with previous literature suggesting that ST conditions may trigger anxiety and stress in vulnerable individuals that subsequently impairs mental processing and task performance (Bosson, Haymovitz & Pinel, 2004; Schmader, Johns & Forbes, 2008; Johns, Inzlicht & Schmader, 2010). Thus, the results of our study suggest that the ST manipulation may have triggered stress and anxiety in those predisposed to these emotions, which in turn, impaired their task performance.

Differences between the ST and NT groups were found in the ERP measures of internal response monitoring as measured via the ERN and CRN components. Although our initial hypothesis was that ST would lead to an increase in monitoring errors, our results suggest an overall enhancement of monitoring both accurate and inaccurate responses. Enhancement in the ERN component alone suggests a specific increase in error-monitoring, (Luu, Flaisch & Tucker, 2004), however, a change in both the ERN and the CRN components is hypothesized to reflect a more general change in response-monitoring (Hajcak, McDonald & Simons, 2004; Moser, Moran & Jendrusina, 2012; Endrass et al., 2008). Importantly, previous work has shown that both the ERN and the CRN are enhanced in anxious individuals (Hajcak, McDonald & Simons, 2004; Endrass et al., 2008), particularly in relation to anxious apprehension (i.e. worry; Moser, Moran & Jendrusina, 2012), linked with corresponding cognitive inefficiency (Endrass et al., 2008; Eysenck et al., 2007). Thus, this pattern of increased ERN and CRN under ST, combined with a lack of relation between ERN and CRN amplitudes and task performance, suggests an increase in anxious apprehension and a subsequent decrease in cognitive efficiency under ST conditions.

Even though the group difference in ERN and CRN amplitudes was seen during easy trials, the relation between math identification and the ERN difference score only appeared during difficult trials. Higher math identification predicted attenuated ERN reactivity, controlling for CRN reactivity, during difficult trials. As attenuated ERN amplitude was not related to an increase in performance, these results are unlikely to represent an increase in error-monitoring efficiency. Instead, these results related to anxiety and stress levels in the ST group, with higher anxiety and stress scores predicting more attenuated ERN difference scores to difficult trials. This pattern suggests that individuals who cared more about their math ability may have been disengaging from error-monitoring under ST conditions during difficult trials. These results provide neural evidence corresponding with the theory that individuals who are more invested in a threatening domain may use more disengagement to protect their self-esteem (Crocker & Major, 1989; Major, Spencer, Schmader, Wolfe & Crocker, 1998; Woodcock, Hernandez, Estrada & Schultz, 2012).

Group differences also emerged for the Pe component. Specifically, the ST group showed a larger $Pe_{(incorrect)}$ peak-to-peak score to easy errors than the NT group. Since the Pe is associated with conscious awareness of an error (Nieuwenhuis et al., 2001; Hajcak, McDonald & Simons, 2003; Hughes & Yeung, 2011), and is thought to increase along with the motivational significance of the error (Overbeek, Nieuwenhuis & Ridderinkhof, 2005; Ullsperger et al., 2007), the increased amplitude among women in the ST group may reflect an increase in both conscious error processing and the saliency attributed to errors. However, the lack of a relation between $Pe_{(incorrect)}$ peak-to-peak amplitudes and task accuracy (i.e., percent correct) in the ST group suggests that participants are not using enhanced attention to errors to improve their performance. Instead, there was a relation between $Pe_{(incorrect)}$ peak-to-peak amplitude and errors of omission, with higher $Pe_{(incorrect)}$ peak-to-peak amplitude predicting more errors of omission in the ST group.

Combined, these data suggest that participants under ST are more consciously attending to errors but are reacting to that information by disengaging from the task, rather than using it to better their performance in an efficient manner. Thus, the Pe findings from Study 2 correspond to behavioral literature on ST indicating a negative bias in error processing under stress that is unrelated to task accuracy (Cadinu et al., 2005; Brozovich & Heimberg, 2008; Morgan & Banerjee, 2008; Cody & Teachman, 2010). These findings highlight the importance of implementing neural measures in the study of ST, as they are able to differentiate performance-monitoring processes at a level beyond that of behavioral studies alone; specifically, here we show that ST impacts performance monitoring prior to the implementation of external feedback – a distinction that occurs too quickly to be measured by behavioral paradigms.

There was no group difference in FRN amplitude suggesting that the ST manipulation did not alter immediate feedback processing among young women. However, there was a group difference in the relation between ERN difference scores and FRN difference scores. Within the NT group, a higher ERN amplitude to easy errors predicted higher FRN amplitude to negative feedback during easy trials. Although the ERN-FRN relation in the current study is the opposite of the pattern typically seen in the literature (Heldmann, Russeler and Munte, 2008; Stahl et al., 2010), this finding may be due to a difference in the timing of the task in the current study and subsequent predictability. For example, the flanker task used in the study by Stahl and colleagues (2010) showed the response stimuli for 900ms, allowing the participants to more consciously determine a correct answer, and thus monitor their responses with more certainty. In the current study, the response stimulus was only shown for 200ms, which may have led to more uncertainty in participant monitoring of responses. Therefore, the women in the NT condition may have been using the external feedback to verify their internal monitoring. This possibility is supported by the relation between the FRN and accuracy rates (i.e., percent correct) in the NT group, with higher FRN amplitudes to negative feedback on easy errors predicting better task performance. The lack of this relation within the ST group, therefore, may suggest a less adaptive use of error-monitoring within this group.

The current study does have some limitations. The basic limitations remain the same as in Study 1, with participants recruited from one discipline, and coming from a community lacking in racial diversity. In Study 2, the sample size was small compared to previous behavioral studies examining ST (Johns et al., 2008; Stout et al., 2011), however, it was within the range of previous ERP studies exploring ST effects (Forbes & Leitner, 2014; Forbes et al., 2008) and the effect sizes for the analysis described above are all within the medium-large range.

Additionally, although there was not a significant relation between neural reactivity to errors and the choice to complete the fourth round, this lack of association may have been influenced by the capping procedure itself. The addition of the capping procedure, with its slight discomfort and the lengthening of the experimental time frame, may inhibit participant's self-motivation to complete the fourth task (as observed in Study 1). Furthermore, in Study 2, participants were more likely to complete the fourth round if they reported lower confidence in their task ability. Previous work shows that lower self-confidence predicts more compliance, particularly among women (Gudjonsson et al., 2002; Gudjonsson & Sigurdsson, 2003). The researcher-to-participant ratio in Study 2, with two research assistants for every participant, may have heightened participants' motivation to comply with the researcher's request to complete the fourth round. Future studies should explore task motivation under other conditions that may decrease these situational stressors (i.e., a fourth round without the EEG cap; an optional round in the middle of the task, rather than at the end) and its potential relations to increased response monitoring under ST.

Despite these limitations, the current study advances our understanding of the impact of ST on women's performance monitoring in several important ways. First, the findings suggest that women under ST are monitoring errors in an inefficient manner, showing enhanced immediate processing of both errors and correct responses, similar to patterns seen in clinically anxious individuals. Secondly, women under ST are consciously focusing on their errors to a greater extent than those in the NT group, to the detriment of their overall performance, suggesting processes such as mal-adaptive rumination. These patterns in response monitoring under ST may help explain the lower task motivation, task disengagement and later domain de-identification associated with long-term ST which has been shown in previous literature.

Table 3a. Means and standard errors for age, anxiety scores, stress scores, math anxiety level and math identification scores in Study 2

	Stereotype Threat Mean *(SE)*	Non-Threat Mean *(SE)*
Final Sample Size	22	18
Age	19.09*(0.20)*	19.72*(0.25)*
Anxiety	9.05*(2.11)*	5.78*(1.40)*
Stress	11.62*(2.34)*	7.11*(1.58)*
Math Anxiety Level	1.91*(0.15)*	1.56*(0.12)*
Math ID	58.00*(1.67)*	54.72*(1.28)*

Table 3b. Means, standard errors and between group comparisons (p-values) for the number of epochs acquired following response onset and feedback onset, controlling for reaction time, time of day, age, and math anxiety.

Trigger Type	Trial Level	Stereotype Threat Mean*(SE)*	Non-Threat Mean*(SE)*	*p*
Error Response	Easy	20.13*(1.84)*	23.90*(2.14)*	.214
	Hard	37.82*(2.41)*	38.49*(2.80)*	.865
Correct Response	Easy	82.85*(5.69)*	92.38*(6.63)*	.310
	Hard	67.40*(4.10)*	78.03*(4.77)*	.120
Negative Feedback	Easy	19.43*(1.79)*	23.62*(2.01)*	.158
	Hard	36.82*(2.28)*	38.11*(2.65)*	.731
Positive Feedback	Easy	86.70*(4.97)*	95.05*(5.78)*	.548
	Hard	66.96*(4.01)*	77.68*(4.67)*	.109

Table 4. Number of participants who chose to complete the optional fourth task round by group in Study 2.

	Stereotype Threat	Non-Threat
Yes to 4^{th} Round	6	8
No to 4^{th} Round	16	10

Table 5a. Regressions between task performance measures and ERN reactivity to easy and hard trials, controlling for age and math anxiety in Study 2.

Group	Performance Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Stereotype Threat						
	Errors of Omission					
		Easy	-5.70 *(4.45)*	-.29	-1.20	.25
		Hard	-9.61 *(5.70)*	-.41	-1.69	.11
	Percent Correct					
		Easy	-1.23 *(0.78)*	-.30	-1.58	.13
		Hard	-0.91 *(1.02)*	-.18	-0.89	.38
	Post-Error Slowing					
		Easy	-5.11 *(6.97)*	-.16	-0.73	.47
		Hard	-4.46 *(8.75)*	-.12	-0.51	.62
Non-Threat						
	Errors of Omission					
		Easy	5.99 *(5.89)*	.27	1.02	.33
		Hard	4.97 *(5.72)*	.24	0.87	.40
	Percent Correct					
		Easy	0.29 *(1.01)*	.07	0.29	.78
		Hard	0.65 *(0.96)*	.16	0.68	.51
	Post-Error Slowing					
		Easy	6.59 *(6.14)*	.29	1.07	.30
		Hard	5.69 *(5.95)*	.26	0.96	.36

*$p<.05$

Table 5b. Regressions between task performance measures and CRN reactivity to easy and hard trials, controlling for age and math anxiety in Study 2.

Group	Performance Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Stereotype Threat						
	Errors of Omission					
		Easy	-7.73 *(4.82)*	-.38	-1.60	.13
		Hard	-9.88 *(6.70)*	-.34	-1.48	.16
	Percent Correct					
		Easy	-0.48 *(0.86)*	-.11	-0.55	.59
		Hard	-2.01 *(1.10)*	-.34	-1.84	.08
	Post-Error Slowing					
		Easy	-6.06 *(7.25)*	-.18	-0.84	.42
		Hard	-13.20 *(9.67)*	-.28	-1.37	.19
Non-Threat						
	Errors of Omission					
		Easy	-0.64 *(5.56)*	-.03	-0.11	.91
		Hard	3.33 *(6.35)*	.15	0.53	.61
	Percent Correct					
		Easy	0.61 *(0.91)*	.15	0.67	.51
		Hard	1.28 *(1.01)*	.28	1.27	.23
	Post-Error Slowing					
		Easy	7.91 *(5.40)*	.36	1.47	.17
		Hard	6.71 *(6.45)*	.28	1.04	.32

*$p < .05$

Table 5c. Regressions between task performance measures and ERN difference scores (ERN minus CRN) to easy and hard trials, controlling for age and math anxiety in Study 2.

Group	Performance Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Stereotype Threat						
	Errors of Omission					
		Easy	1.13 *(4.47)*	.07	0.25	.80
		Hard	-6.24 *(10.27)*	-.18	-0.61	.55
	Percent Correct					
		Easy	-0.65 *(0.74)*	-.18	-0.88	.39
		Hard	1.71 *(1.70)*	.24	1.00	.33
	Post-Error Slowing					
		Easy	0.36 *(6.42)*	.01	0.06	.96
		Hard	15.46 *(14.38)*	.28	1.08	.30
Non-Threat						
	Errors of Omission					
		Easy	7.36 *(6.06)*	.33	1.22	.25
		Hard	5.63 *(9.34)*	.16	0.60	.56
	Percent Correct					
		Easy	-0.48 *(1.05)*	-.11	-0.46	.65
		Hard	-1.11 *(1.54)*	-.16	-0.72	.49
	Post-Error Slowing					
		Easy	-3.25 *(6.63)*	.14	-0.49	.63
		Hard	0.11 *(9.91)*	.003	0.01	.99

*$p<.05$

Table 6a. Regressions between participant attitude measures and ERN reactivity to easy and hard trials, controlling for age and percent correct in Study 2.

Group	Attitude Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Stereotype Threat						
	Math ID					
		Easy	-.95 *(1.24)*	-.21	-0.77	.46
		Hard	.64 *(1.73)*	.11	0.37	.72
	General Anxiety					
		Easy	-.85 *(1.08)*	-.14	-0.78	.45
		Hard	.61 *(1.28)*	.08	0.48	.64
	General Stress					
		Easy	-.10 *(1.39)*	-.01	-0.07	.94
		Hard	2.04 *(1.54)*	.25	1.33	.20
	Task Confidence					
		Easy	-.13 *(0.09)*	-.37	-1.47	.16
		Hard	-.02 *(0.11)*	-.05	-0.19	.85
Non-Threat						
	Math ID					
		Easy	1.27 *(0.96)*	.32	1.34	.20
		Hard	-.33 *(0.95)*	-.09	-0.35	.73
	General Anxiety					
		Easy	1.64 *(1.15)*	.37	1.43	.18
		Hard	-.86 *(1.13)*	-.21	-0.76	.46
	General Stress					
		Easy	1.90 *(1.27)*	.39	1.50	.16
		Hard	-.20 *(1.28)*	-.04	-0.16	.88
	Task Confidence					
		Easy	.01 *(0.13)*	.03	0.10	.92
		Hard	.11 *(0.10)*	.31	1.10	.29

*$p<.05$

Table 6b. Regressions between participant attitude measures and CRN reactivity to easy and hard trials, controlling for age and percent correct in Study 2.

Group	Attitude Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Stereotype Threat						
	Math ID					
		Easy	-1.09 *(1.29)*	.23	-0.85	.41
		Hard	-3.06 *(1.75)*	-.45	-1.75	.10
	General Anxiety					
		Easy	-0.21 *(1.07)*	-.03	-0.20	.85
		Hard	-2.04 *(1.43)*	-.23	-1.42	.17
	General Stress					
		Easy	0.67 *(1.34)*	-.10	-0.50	.62
		Hard	-1.60 *(1.88)*	-.16	-0.86	.41
	Task Confidence					
		Easy	-0.004*(.09)*	-.01	-0.04	.97
		Hard	-0.10 *(.13)*	-.20	-0.78	.44
Non-Threat						
	Math ID					
		Easy	-0.60 *(0.91)*	-.16	-0.66	.52
		Hard	-0.09 *(1.06)*	-.02	-0.08	.94
	General Anxiety					
		Easy	-0.15 *(1.13)*	-.04	-0.13	.90
		Hard	-0.17 *(1.29)*	-.04	-0.13	.90
	General Stress					
		Easy	0.24 *(1.24)*	.05	0.19	.85
		Hard	0.29 *(1.42)*	.06	0.20	.84
	Task Confidence					
		Easy	-0.04 *(0.10)*	-.12	-0.43	.68
		Hard	0.02 *(0.12)*	.05	0.15	.88

**p*<.05

Table 6c. Regressions between participant attitude measures and ERN difference scores (ERN minus CRN) on easy and hard trials, controlling for age and percent correct in Study 2.

Group	Attitude Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Stereotype Threat						
	Math ID					
		Easy	0.04 *(1.18)*	.01	0.04	.97
		Hard	4.36 *(1.71)*	.55	2.55	.02*
	General Anxiety					
		Easy	-0.46 *(0.94)*	-.09	-0.49	.63
		Hard	3.74 *(1.45)*	.35	1.57	.02*
	General Stress					
		Easy	-0.59 *(1.18)*	-.10	-0.50	.62
		Hard	5.77 *(1.63)*	.48	3.54	.003*
	Task Confidence					
		Easy	-0.10 *(0.08)*	-.31	-1.20	.25
		Hard	0.09 *(0.15)*	.15	0.62	.54
Non-Threat						
	Math ID					
		Easy	2.15 *(0.87)*	.53	2.46	.03*
		Hard	-0.75 *(1.60)*	-.12	-0.47	.65
	General Anxiety					
		Easy	1.95 *(1.16)*	.44	1.67	.12
		Hard	-2.05 *(1.87)*	-.29	-1.10	.29
	General Stress					
		Easy	1.72 *(1.33)*	.35	1.29	.22
		Hard	-1.22 *(2.13)*	-.16	-0.57	.58
	Task Confidence					
		Easy	0.09 *(0.13)*	.22	0.67	.52
		Hard	0.27 *(0.16)*	.46	1.72	.11

p<.05

Table 7a. Regressions between task performance measures and $Pe_{(incorrect)}$ reactivity to easy and hard trials, controlling for age and math anxiety in Study 2.

Group	Performance Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Stereotype Threat						
	Errors of Omission					
		Easy	7.25 *(3.23)*	.51	2.25	.04*
		Hard	9.40 *(3.50)*	.60	2.68	.02*
	Percent Correct					
		Easy	-0.61 *(0.61)*	-.20	-0.99	.34
		Hard	-1.72 *(0.57)*	-.52	-3.01	.01*
	Post-Error Slowing					
		Easy	-6.48 *(5.17)*	-.27	-1.25	.23
		Hard	-13.40 *(5.15)*	-.52	-2.60	.02*
Non-Threat						
	Errors of Omission					
		Easy	-3.63 *(2.06)*	-.44	-1.76	.10
		Hard	-0.89 *(5.52)*	-.05	-0.16	.88
	Percent Correct					
		Easy	0.10 *(0.43)*	.05	0.23	.83
		Hard	0.68 *(1.02)*	.17	0.66	.52
	Post-Error Slowing					
		Easy	-0.15 *(2.90)*	-.02	-0.05	.96
		Hard	1.01 *(6.91)*	.05	0.15	.89

*$p<.05$

Table 7b. Regressions between task performance measures and $Pe_{(correct)}$ reactivity to easy and hard trials, controlling for age and math anxiety in Study 2.

Group	Performance Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Stereotype Threat						
	Errors of Omission					
		Easy	9.56 *(3.35)*	.61	2.86	.01*
		Hard	10.19 *(4.00)*	.58	2.55	.02*
	Percent Correct					
		Easy	-1.75 *(0.54)*	-.53	-3.25	.01*
		Hard	-2.05 *(0.62)*	-.55	-3.33	.01*
	Post-Error Slowing					
		Easy	-13.94 *(4.87)*	-.54	-2.86	.01*
		Hard	-17.11 *(5.40)*	-.59	-3.17	.01*
Non-Threat						
	Errors of Omission					
		Easy	-1.05 *(1.14)*	-.23	-.91	.38
		Hard	-1.05 *(9.56)*	-.03	-.11	.92
	Percent Correct					
		Easy	-0.05 *(0.22)*	-.05	-.22	.83
		Hard	-1.03 *(1.77)*	-.15	-.58	.57
	Post-Error Slowing					
		Easy	-0.20 *(1.48)*	-.04	-.13	.90
		Hard	3.09 *(11.94)*	.09	.26	.80

*$p<.05$

Table 7c. Regressions between task performance measures and Pe difference scores ($Pe_{(incorrect)}$ minus $Pe_{(correct)}$) on easy and hard trials, controlling for age and math anxiety in Study 2.

Group	Performance Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Stereotype Threat						
	Errors of Omission					
		Easy	0.90 *(5.00)*	.05	0.18	.86
		Hard	8.65 *(12.73)*	.19	0.68	.51
	Percent Correct					
		Easy	0.57 *(0.70)*	.19	0.82	.43
		Hard	0.75 *(1.83)*	.09	0.41	.69
	Post-Error Slowing					
		Easy	-0.65 *(5.26)*	-.03	-0.12	.90
		Hard	18.38 *(12.68)*	.34	1.45	.17
Non-Threat						
	Errors of Omission					
		Easy	0.28 *(2.01)*	.04	0.14	.89
		Hard	-0.64 *(6.05)*	-.03	-0.11	.92
	Percent Correct					
		Easy	0.21 *(0.37)*	.12	0.57	.58
		Hard	1.23 *(1.08)*	.25	1.13	.28
	Post-Error Slowing					
		Easy	0.45 *(2.51)*	.05	0.18	.86
		Hard	-0.03 *(7.58)*	-.001	-0.004	.99

*$p<.05$

Table 8a. Regressions between participant attitude measures and $Pe_{(incorrect)}$ reactivity to easy and hard trials, controlling for age and percent correct in Study 2.

Group	Attitude Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Stereotype Threat						
	Math ID					
		Easy	-1.19 *(0.61)*	-.42	-1.96	.07
		Hard	-0.02 *(1.05)*	-.004	-0.02	.99
	General Anxiety					
		Easy	-0.77 *(0.73)*	-.19	-1.06	.31
		Hard	-1.03 *(1.12)*	-.17	-0.92	.37
	General Stress					
		Easy	-1.99 *(0.89)*	-.39	-2.24	.04*
		Hard	-2.87 *(1.38)*	-.38	-2.08	.06
	Task Confidence					
		Easy	0.02 *(0.07)*	.08	0.30	.77
		Hard	0.10 *(0.12)*	.23	0.83	.42
Non-Threat						
	Math ID					
		Easy	-0.53 *(0.40)*	-.36	-1.34	.21
		Hard	-0.38 *(1.03)*	-.12	-0.37	.72
	General Anxiety					
		Easy	-0.69 *(0.50)*	-.39	-1.38	.19
		Hard	0.87 *(1.28)*	.22	0.68	.51
	General Stress					
		Easy	-0.92 *(0.53)*	-.45	-1.75	.11
		Hard	-1.42 *(1.36)*	-.32	-1.04	.32
	Task Confidence					
		Easy	0.01 *(0.05)*	.04	0.12	.91
		Hard	-0.001*(0.12)*	-.002	-0.01	.99

*$p<.05$

Table 8b. Regressions between participant attitude measures and $Pe_{(correct)}$ to easy and hard trials, controlling for age and percent correct in Study 2.

Group	Attitude Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Stereotype Threat						
	Math ID					
		Easy	-0.17 *(1.14)*	-.04	-0.15	.88
		Hard	-0.51 *(1.23)*	-.11	-0.42	.68
	General Anxiety					
		Easy	-2.82 *(1.53)*	-.35	-1.84	.09
		Hard	-1.49 *(1.27)*	-.23	-1.17	.26
	General Stress					
		Easy	-0.82 *(1.14)*	-.15	-0.72	.48
		Hard	-3.01 *(1.65)*	-.37	-1.83	.09
	Task Confidence					
		Easy	0.09 *(.12)*	.20	0.76	.46
		Hard	0.02 *(.14)*	.04	0.13	.90
Non-Threat						
	Math ID					
		Easy	-0.14 *(0.21)*	-.17	-0.64	.54
		Hard	-0.05 *(1.72)*	-.01	-0.03	.98
	General Anxiety					
		Easy	-0.28 *(0.26)*	-.29	-1.06	.31
		Hard	-0.65 *(2.17)*	-.10	-0.30	.77
	General Stress					
		Easy	-0.32 *(0.29)*	-.29	-1.12	.28
		Hard	-1.09 *(2.36)*	-.14	-0.46	.65
	Task Confidence					
		Easy	0.004 *(0.03)*	.05	0.15	.88
		Hard	0.23 *(0.20)*	.37	1.14	.28

*$p<.05$

Table 8c. Regressions between participant attitude measures and Pe difference scores ($Pe_{(incorrect)}$ minus $Pe_{(correct)}$) on easy and hard trials, controlling for age and percent correct in Study 2.

Group	Attitude Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Stereotype Threat						
	Math ID					
		Easy	-1.79 *(0.71)*	-.52	-2.51	.03*
		Hard	1.51 *(2.12)*	.17	.71	.49
	General Anxiety					
		Easy	-0.96 *(0.92)*	-.20	-1.04	.32
		Hard	0.31 *(2.30)*	.03	0.13	.90
	General Stress					
		Easy	-1.52 *(1.26)*	-.25	-1.21	.25
		Hard	-2.55 *(3.10)*	-.17	-0.82	.43
	Task Confidence					
		Easy	-0.02 *(0.10)*	-.05	-0.17	.87
		Hard	0.33 *(0.22)*	.38	1.49	.16
Non-Threat						
	Math ID					
		Easy	-0.01 *(0.37)*	-.01	-0.02	.98
		Hard	-0.44 *(1.13)*	-.11	-0.39	.70
	General Anxiety					
		Easy	0.29 *(0.46)*	.17	0.61	.55
		Hard	1.33 *(1.39)*	.28	0.96	.36
	General Stress					
		Easy	0.24 *(0.51)*	.13	0.47	.65
		Hard	-1.25 *(1.53)*	-.23	-0.82	.43
	Task Confidence					
		Easy	-0.01 *(0.04)*	-.05	-0.16	.88
		Hard	-0.09 *(0.13)*	-.23	-0.71	.49

*$p<.05$

Table 9a. Regressions between task performance measures and $FRN_{(negative)}$ reactivity to easy and hard trials, controlling for age and math anxiety in Study 2.

Group	Performance Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Stereotype Threat						
	Errors of Omission					
		Easy	0.15 *(3.86)*	.01	0.04	.97
		Hard	-6.64 *(5.03)*	-.32	-1.32	.20
	Percent Correct					
		Easy	-0.80 *(0.62)*	-.26	-1.29	.21
		Hard	0.21 *(0.89)*	.05	0.23	.82
	Post-Error Slowing					
		Easy	-2.95 *(5.48)*	-.12	-0.54	.60
		Hard	3.95 *(7.49)*	.12	0.53	.61
Non-Threat						
	Errors of Omission					
		Easy	-2.49 *(3.52)*	-.21	-0.71	.49
		Hard	-1.42 *(3.23)*	-.12	-0.44	.67
	Percent Correct					
		Easy	-0.62 *(0.57)*	-.26	-1.08	.30
		Hard	-0.45 *(0.53)*	-.19	-0.86	.41
	Post-Error Slowing					
		Easy	2.68 *(3.68)*	.21	0.73	.48
		Hard	-0.04 *(3.41)*	-.004	-0.01	.99

*$p<.05$

Table 9b. Regressions between task performance measures and $FRN_{(positive)}$ reactivity to easy and hard trials, controlling for age and math anxiety in Study 2.

Group	Performance Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Stereotype Threat						
	Errors of Omission					
		Easy	4.09 *(4.39)*	.24	0.93	.36
		Hard	8.77 *(5.84)*	.38	1.50	.15
	Percent Correct					
		Easy	0.36 *(0.75)*	.10	0.48	.64
		Hard	0.20 *(1.05)*	.04	0.19	.85
	Post-Error Slowing					
		Easy	1.89 *(6.42)*	.07	0.29	.77
		Hard	1.37 *(8.88)*	.04	0.15	.88
Non-Threat						
	Errors of Omission					
		Easy	2.70 *(3.66)*	.20	0.74	.47
		Hard	0.99 *(3.91)*	.07	0.25	.80
	Percent Correct					
		Easy	0.26 *(0.62)*	.10	0.42	.68
		Hard	-0.28 *(0.65)*	-.10	-0.43	.68
	Post-Error Slowing					
		Easy	2.76 *(3.84)*	.20	0.72	.48
		Hard	4.35 *(3.92)*	.29	1.11	.29

*$p<.05$

Table 9c. Regressions between task performance measures and FRN difference scores ($FRN_{(negative)}$ minus $FRN_{(positive)}$) on easy and hard trials, controlling for age and math anxiety in Study 2.

Group	Performance Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Stereotype Threat						
	Errors of Omission					
		Easy	-3.75 *(4.32)*	-.21	-0.87	.40
		Hard	-13.20 *(4.26)*	-.61	-3.10	.006*
	Percent Correct					
		Easy	-1.40 *(0.66)*	-.38	-2.10	.05
		Hard	0.06 *(0.90)*	.01	0.07	.94
	Post-Error Slowing					
		Easy	-5.68 *(6.17)*	-.20	-0.92	.37
		Hard	3.01 *(7.59)*	.08	0.40	.70
Non-Threat						
	Errors of Omission					
		Easy	-6.69 *(3.72)*	-.45	-1.80	.10
		Hard	-5.87 *(5.19)*	-.29	-1.13	.28
	Percent Correct					
		Easy	1.15 *(0.61)*	-.39	-1.88	.08
		Hard	-0.73 *(0.88)*	-.19	-0.83	.42
	Post-Error Slowing					
		Easy	0.18 *(4.35)*	.01	0.04	.97
		Hard	-8.49 *(5.18)*	-.41	-1.64	.13

*$p<.05$

Table 10a. Regressions between participant attitude measures and $FRN_{(negative)}$ reactivity to easy and hard trials, controlling for age and percent correct in Study 2.

Group	Attitude Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Stereotype Threat						
	Math ID					
		Easy	-0.01 *(1.03)*	-.004	-0.01	.99
		Hard	-0.03 *(1.38)*	-.01	-0.02	.99
	General Anxiety					
		Easy	-1.03 *(0.84)*	-.20	-1.22	.24
		Hard	-1.04 *(1.18)*	-.15	-0.88	.39
	General Stress					
		Easy	-0.20 *(1.11)*	-.04	-0.18	.86
		Hard	-0.07 *(1.52)*	-.01	-0.05	.96
	Task Confidence					
		Easy	0.12 *(0.07)*	.45	1.85	.08
		Hard	0.04 *(0.10)*	.11	0.45	.66
Non-Threat						
	Math ID					
		Easy	0.59 *(0.60)*	.27	0.99	.34
		Hard	0.33 *(0.55)*	.15	0.60	.56
	General Anxiety					
		Easy	0.57 *(0.73)*	.23	0.78	.45
		Hard	0.22 *(0.67)*	.09	0.33	.75
	General Stress					
		Easy	0.87 *(0.79)*	.32	1.11	.29
		Hard	0.53 *(0.73)*	.20	0.72	.49
	Task Confidence					
		Easy	-0.03 *(0.07)*	-.17	-0.52	.61
		Hard	0.01 *(0.06)*	.04	0.13	.90

*$p<.05$

Table 10b. Regressions between participant attitude measures and $FRN_{(positive)}$ reactivity to easy and hard trials, controlling for age and percent correct in Study 2.

Group	Attitude Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Stereotype Threat						
	Math ID					
		Easy	0.84 *(1.04)*	.21	0.81	.43
		Hard	1.51 *(1.45)*	.28	1.05	.31
	General Anxiety					
		Easy	0.65 *(0.90)*	.12	0.72	.48
		Hard	0.56 *(1.28)*	.08	0.44	.67
	General Stress					
		Easy	0.88 *(1.14)*	.14	0.78	.45
		Hard	0.09 *(1.63)*	.01	0.06	.96
	Task Confidence					
		Easy	0.16 *(0.07)*	.49	2.25	.04*
		Hard	0.23 *(0.10)*	.54	2.33	.03*
Non-Threat						
	Math ID					
		Easy	0.02 *(0.60)*	.01	0.04	.97
		Hard	0.31 *(0.65)*	.12	0.48	.64
	General Anxiety					
		Easy	-0.28 *(0.72)*	-.10	-0.39	.71
		Hard	0.19 *(0.80)*	.07	0.24	.82
	General Stress					
		Easy	0.07 *(0.80)*	.03	0.09	.93
		Hard	0.91 *(0.85)*	.29	1.08	.30
	Task Confidence					
		Easy	0.06 *(0.06)*	.25	0.87	.40
		Hard	0.02 *(0.07)*	.06	0.21	.84

*$p<.05$

Table 10c. Regressions between participant attitude measures and FRN differences scores ($FRN_{(negative)}$ minus $FRN_{(positive)}$) on easy and hard trials, controlling for age and percent correct in Study 2.

Group	Attitude Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Stereotype Threat						
	Math ID					
		Easy	-1.09 *(1.16)*	-.26	-0.94	.36
		Hard	-1.27 *(1.32)*	-.24	-0.96	.35
	General Anxiety					
		Easy	-2.26 *(0.87)*	-.40	-2.60	.02*
		Hard	-1.43 *(1.11)*	-.20	-1.29	.22
	General Stress					
		Easy	-1.41 *(1.26)*	-.22	-1.12	.28
		Hard	-0.14 *(1.47)*	-.02	-0.10	.92
	Task Confidence					
		Easy	-0.01 *(0.09)*	-.04	-0.16	.88
		Hard	-0.12 *(0.09)*	-.31	-1.36	.19
Non-Threat						
	Math ID					
		Easy	0.88 *(0.73)*	.32	1.20	.25
		Hard	0.29 *(0.92)*	.08	0.32	.76
	General Anxiety					
		Easy	1.34 *(0.85)*	.45	1.57	.14
		Hard	0.23 *(1.12)*	.06	0.21	.84
	General Stress					
		Easy	1.23 *(0.97)*	.37	1.27	.23
		Hard	-0.35 *(1.24)*	-.08	-0.28	.78
	Task Confidence					
		Easy	-0.14 *(0.07)*	-.56	-1.95	.08
		Hard	-0.01 *(0.10)*	-.02	-0.07	.94

*$p<.05$

Table 11a. Regressions between task performance measures and $P300_{(negative)}$ reactivity to easy and hard trials, controlling for age and math anxiety in Study 2.

Group	Performance Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Stereotype Threat						
	Errors of Omission					
		Easy	0.05 *(1.84)*	.01	0.03	.98
		Hard	-0.38 *(2.50)*	-.04	-0.15	.88
	Percent Correct					
		Easy	-0.13 *(0.31)*	-.09	-0.43	.68
		Hard	-0.19 *(0.42)*	-.09	-0.45	.66
	Post-Error Slowing					
		Easy	-2.28 *(2.62)*	-.19	-0.87	.40
		Hard	-2.97 *(3.55)*	-.19	-0.83	.42
Non-Threat						
	Errors of Omission					
		Easy	-0.78 *(0.68)*	-.29	-1.14	.28
		Hard	-0.45 *(1.78)*	-.07	-0.25	.81
	Percent Correct					
		Easy	0.04 *(0.14)*	.06	0.27	.80
		Hard	0.27 *(0.33)*	.19	0.81	.43
	Post-Error Slowing					
		Easy	-0.06 *(0.90)*	-.02	-0.07	.95
		Hard	0.95 *(2.22)*	.13	0.43	.68

*$p<.05$

Table 11b. Regressions between task performance measures and $P300_{(positive)}$ reactivity to easy and hard trials, controlling for age and math anxiety in Study 2.

Group	Performance Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Stereotype Threat						
	Errors of Omission					
		Easy	-2.30 *(3.86)*	-.17	-0.60	.56
		Hard	0.93 *(2.86)*	.09	0.33	.75
	Percent Correct					
		Easy	-0.92 *(0.62)*	-.33	-1.49	.16
		Hard	-0.14 *(0.76)*	-.05	-0.19	.85
	Post-Error Slowing					
		Easy	-7.97 *(5.29)*	-.36	-1.51	.15
		Hard	3.99 *(5.52)*	.19	0.72	.48
Non-Threat						
	Errors of Omission					
		Easy	-0.71 *(0.72)*	-.25	-0.98	.34
		Hard	-2.72 *(4.24)*	-.21	-0.64	.53
	Percent Correct					
		Easy	-0.02 *(0.14)*	-.03	-0.11	.91
		Hard	0.34 *(0.80)*	.12	0.43	.68
	Post-Error Slowing					
		Easy	-0.08 *(0.94)*	-.02	-0.08	.93
		Hard	-0.18 *(5.40)*	-.01	-0.03	.97

*$p<.05$

Table 11c. Regressions between task performance measures and P300 difference scores ($P300_{(negative)}$ minus $P300_{(positive)}$) on easy and hard trials, controlling for age and math anxiety in Study 2.

Group	Performance Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Non-Threat						
	Errors of Omission					
		Easy	-1.40 *(2.40)*	-.15	-0.58	.57
		Hard	0.04 *(2.78)*	.004	0.02	.99
	Percent Correct					
		Easy	0.58 *(0.43)*	.28	1.36	.20
		Hard	0.50 *(0.50)*	.21	0.99	.34
	Post-Error Slowing					
		Easy	0.13 *(3.05)*	.01	0.04	.97
		Hard	2.36 *(3.41)*	.19	0.69	.50
Stereotype Threat						
	Errors of Omission					
		Easy	1.20 *(2.66)*	.13	0.45	.66
		Hard	-3.82 *(4.58)*	-.22	-0.83	.42
	Percent Correct					
		Easy	0.16 *(0.45)*	.08	0.34	.74
		Hard	1.84 *(0.64)*	.51	2.90	.01*
	Post-Error Slowing					
		Easy	-1.07 *(3.89)*	-.07	-0.27	.79
		Hard	11.89 *(6.09)*	.42	1.95	.07

*$p<.05$

Table 12a. Regressions between participant attitude measures and $P300_{(negative)}$ reactivity to easy and hard trials, controlling for age and math anxiety in Study 2.

Group	Attitude Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Stereotype Threat						
	Math ID					
		Easy	-0.35 *(0.44)*	-.22	-0.80	.44
		Hard	-0.43 *(0.61)*	-.19	-0.70	.50
	General Anxiety					
		Easy	-0.65 *(0.44)*	-.29	-1.47	.17
		Hard	-0.57 *(0.60)*	-.19	-0.96	.36
	General Stress					
		Easy	-0.52 *(0.64)*	-.19	-0.81	.43
		Hard	-0.20 *(0.86)*	-.05	-0.23	.82
	Task Confidence					
		Easy	-0.05 *(0.05)*	-.29	-0.97	.35
		Hard	-0.07 *(0.06)*	-.31	-1.12	.28
Non-Threat						
	Math ID					
		Easy	-0.10 *(0.13)*	-.21	-.80	.44
		Hard	0.07 *(0.33)*	.06	0.20	.85
	General Anxiety					
		Easy	-0.18 *(0.16)*	-.31	-1.16	.27
		Hard	-0.05 *(0.42)*	-.04	-0.13	.90
	General Stress					
		Easy	-0.29 *(0.16)*	-.43	-1.74	.11
		Hard	-0.40 *(0.44)*	-.25	-0.91	.38
	Task Confidence					
		Easy	0.01 *(0.02)*	.11	.36	.73
		Hard	0.04 *(0.04)*	.35	1.12	.29

*$p<.05$

Table 12b. Regressions between participant attitude measures and P300$_{(positive)}$ reactivity to easy and hard trials, controlling for age and math anxiety in Study 2.

Group	Attitude Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Stereotype Threat						
	Math ID					
		Easy	-1.09 *(0.92)*	-.29	-1.19	.26
		Hard	-1.72 *(0.75)*	-.49	-2.30	.04*
	General Anxiety					
		Easy	-1.09 *(1.01)*	-.21	-1.08	.30
		Hard	-1.59 *(0.79)*	-.33	-2.01	.07
	General Stress					
		Easy	-0.80 *(1.44)*	-.12	-0.55	.59
		Hard	-1.67 *(1.16)*	-.28	-1.44	.17
	Task Confidence					
		Easy	-0.21 *(0.10)*	-.58	-2.24	.04*
		Hard	-0.15 *(0.09)*	-.43	-1.68	.12
Non-Threat						
	Math ID					
		Easy	-0.09 *(0.14)*	-.17	-0.65	.53
		Hard	-0.04 *(0.72)*	-.02	-0.05	.96
	General Anxiety					
		Easy	-0.20 *(0.17)*	-.32	-1.20	.25
		Hard	-0.66 *(0.89)*	-.23	-0.74	.47
	General Stress					
		Easy	-0.24 *(0.18)*	-.35	-1.37	.20
		Hard	-1.49 *(0.89)*	-.46	-1.67	.12
	Task Confidence					
		Easy	.002 *(0.02)*	.05	0.15	.88
		Hard	0.07 *(0.09)*	.29	0.82	.43

*$p<.05$

Table 12c. Regressions between participant attitude measures and P300 difference scores ($P300_{(negative)}$ minus $P300_{(positive)}$) on easy and hard trials, controlling for age and math anxiety in Study 2.

Group	Attitude Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Stereotype Threat						
	Math ID					
		Easy	-0.83 *(0.50)*	-.44	-1.67	.12
		Hard	0.46 *(1.13)*	.14	0.41	.69
	General Anxiety					
		Easy	-0.37 *(0.46)*	-.13	-0.80	.44
		Hard	0.52 *(0.98)*	.10	0.53	.60
	General Stress					
		Easy	-0.49 *(0.64)*	-.15	-0.77	.45
		Hard	1.51 *(1.31)*	.25	1.16	.27
	Task Confidence					
		Easy	-0.01 *(0.05)*	-.07	-0.26	.80
		Hard	-0.01 *(0.10)*	-.03	-0.08	.94
Non-Threat						
	Math ID					
		Easy	-0.28 *(0.47)*	-.17	-0.60	.56
		Hard	0.19 *(0.53)*	.10	0.36	.73
	General Anxiety					
		Easy	-0.02 *(0.60)*	-.01	-0.03	.98
		Hard	0.23 *(0.67)*	.10	0.34	.74
	General Stress					
		Easy	-0.78 *(0.62)*	-.34	-1.26	.23
		Hard	-0.22 *(0.73)*	-.08	-0.30	.77
	Task Confidence					
		Easy	0.04 *(0.06)*	.25	0.77	.46
		Hard	0.07 *(0.06)*	.35	1.18	.26

*$p<.05$

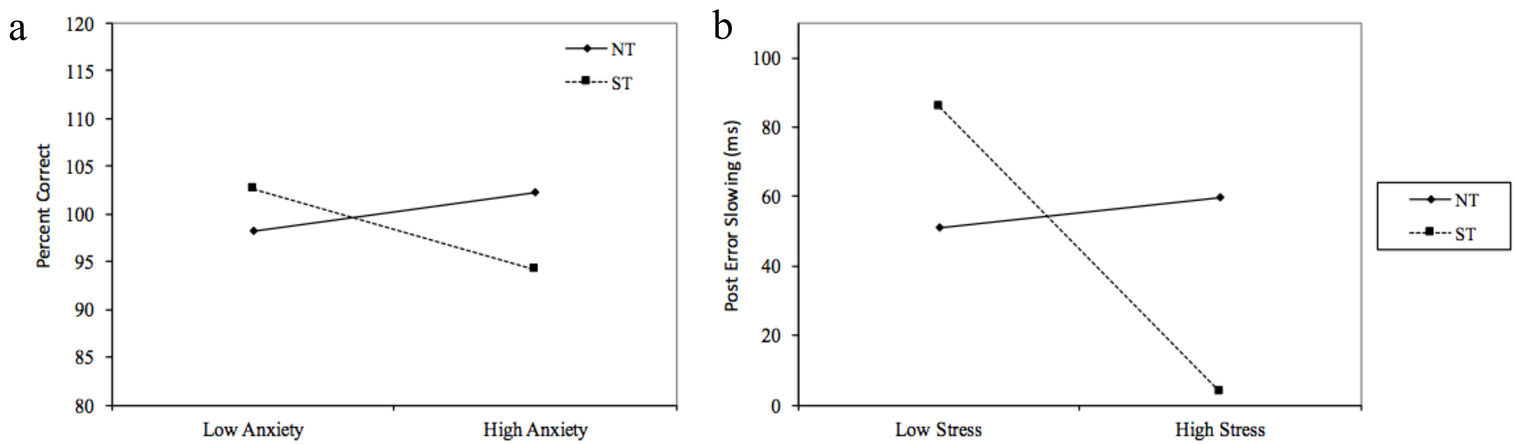


Figure 6. Stereotype threat impact on performance in Study 2. a) Predicted relation between percent correct, condition and general anxiety, controlling for age and math anxiety. b) Predicted relation between post-error slowing, condition and general stress, controlling for age and math anxiety.

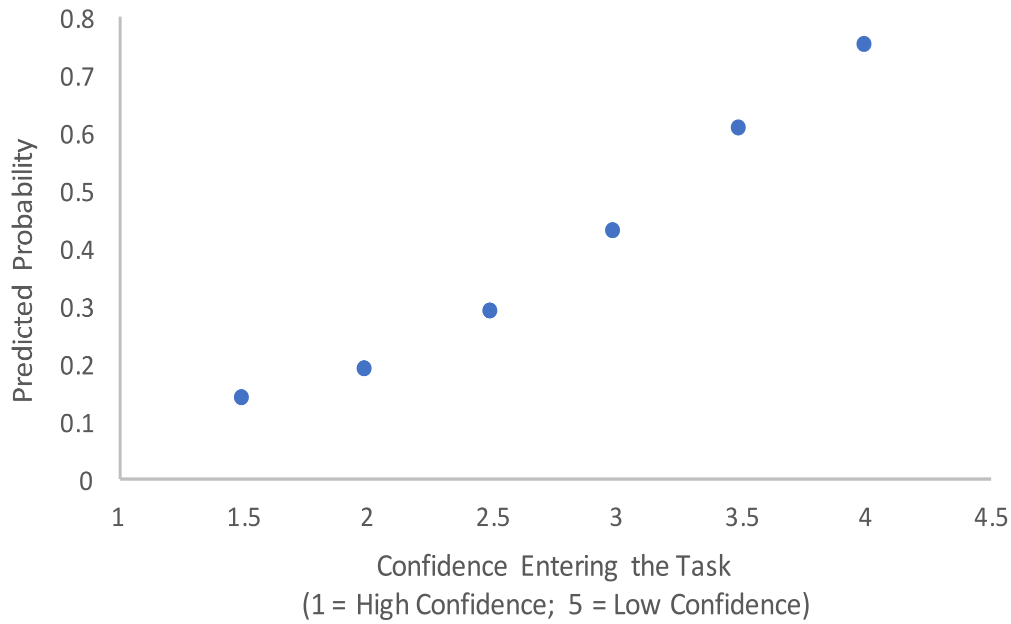


Figure 7. Logistic regression using confidence scores, controlling for percent correct. The patterns show predicted probability of choosing to complete the 4th round.

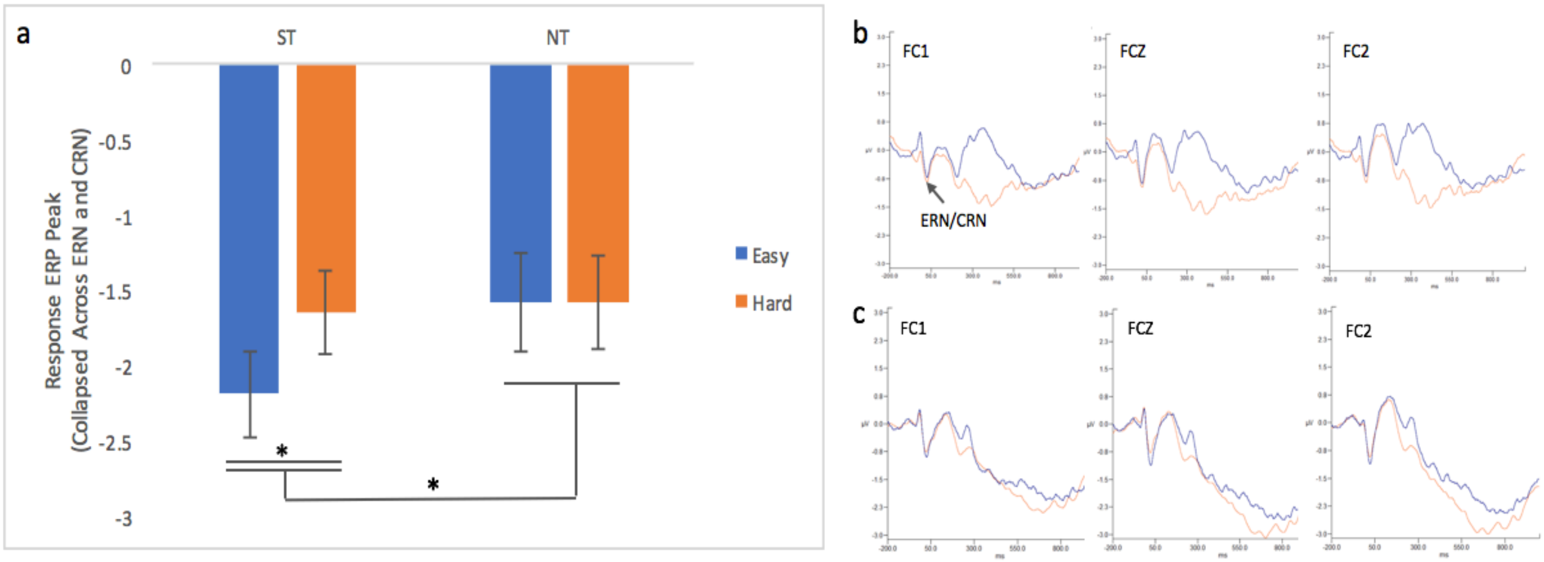


Figure 8. ERP peaks at the frontal-central region (FCZ, FC1 and FC2), averaging across the CRN and ERN. a) Averaged ERN and CRN values, at frontal-central region (FC1, FCZ and FC2), controlling for age, percent correct and math anxiety. Blue bars represent response peaks to easy trials; orange bars represent response peaks to hard trials; error bars represent standard errors. Stars (*) represent significant differences. b) Collapsed ERN and CRN waves in the NT group during easy (blue waves) and hard (orange waves) trials at FC1, FCZ and FC2 sites. ERN and CRN values used for analysis were scored as the most negative peak within 0-100ms post response. c) Collapsed ERN and CRN waves for the ST group during easy (blue waves) and hard (orange waves) trials at FC1, FCZ and FC2 sites. ERN and CRN values used for analysis were scored as the most negative point between 0-100ms post response.

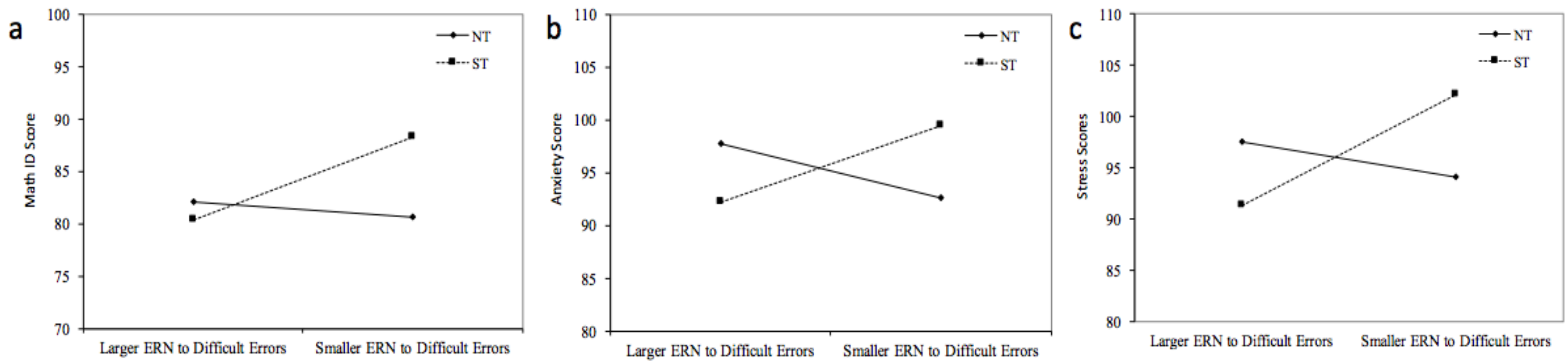


Figure 9. Predicted relation between ERN minus CRN reactivity to difficult errors and a) math identification scores, b) anxiety scores and c) stress scores. Models included age, and percent correct as controls.

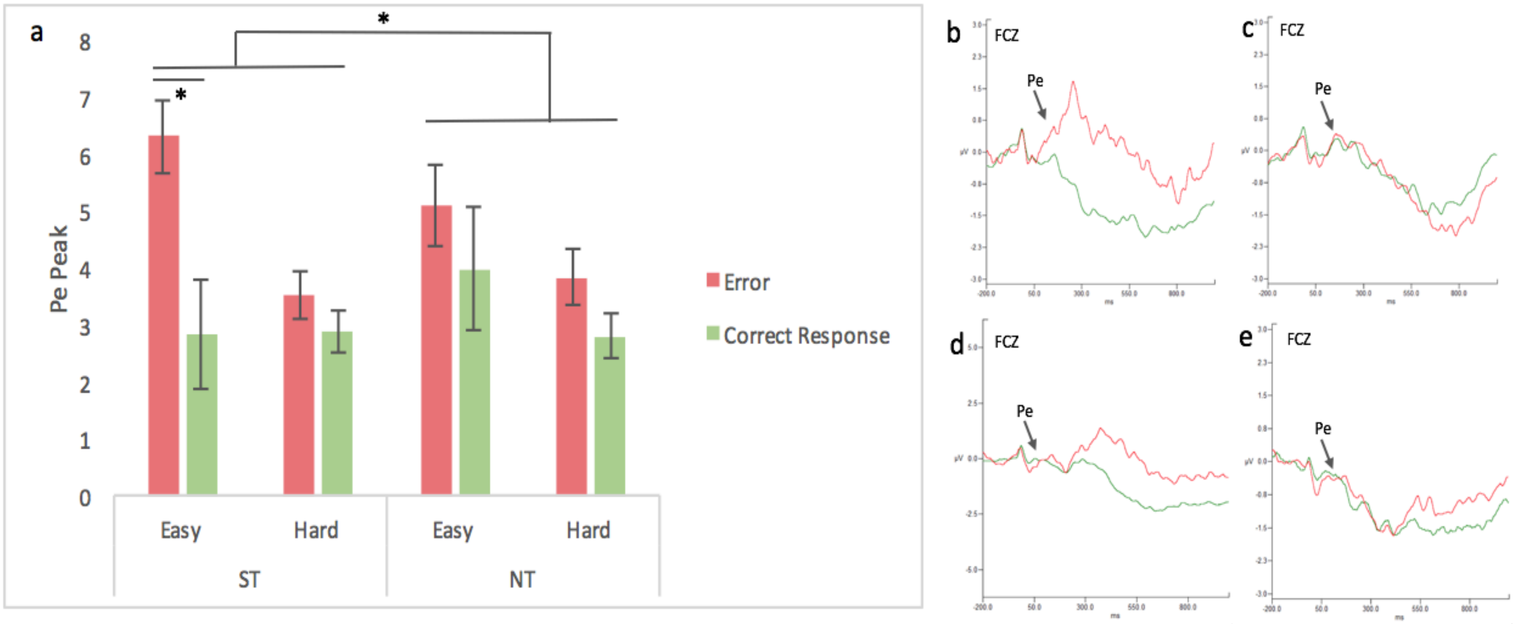


Figure 10. Pe peaks at central-parietal region (Pz, P1 and P2). a) Pe peak values, averaged across sites P1, PZ and P2, controlling for age, percent correct and math anxiety. Green bars represent Pe peaks to correct responses; red bars represent Pe peaks to hard trials; error bars represent standard errors; stars (*) represent significant differences. b) Average Pe waves in the ST group during easy trials at site PZ. Red waves are in response to errors; green waves are in response to correct responses. Pe values used for analysis were scored as the most positive peak within 50-250ms post response. c) Average Pe waves in the ST group during hard trials at site PZ. d) Average Pe waves in the NT group during easy trials at site PZ. e) Average Pe waves in the NT group during hard trials at site PZ.

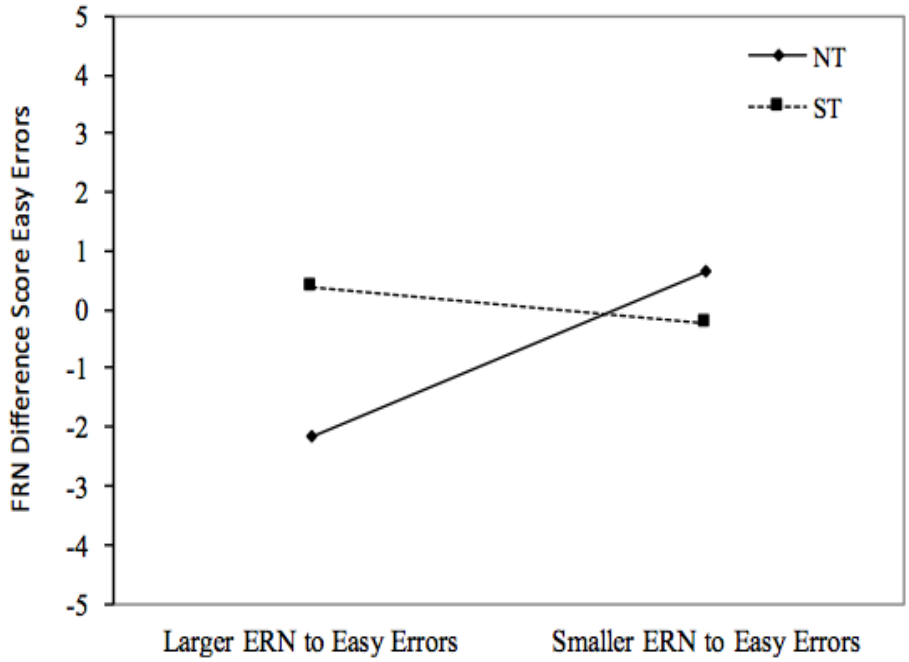


Figure 11. Predicted relation between ERN difference score reactivity to easy errors and FRN difference score reactivity to easy errors, controlling for age, percent correct and math anxiety.

CHAPTER 4

STUDY 3: PERFORMANCE MONITORING IN THE SIM

4.1 Aims and Hypothesis

The aim of Study 3 was to examine whether the Stereotype Inoculation Model (SIM) counters the effects of ST on performance monitoring in young women. We hypothesized that women in a stereotype inoculation (SI) group who read about female STEM experts would attribute less saliency to their errors as evidenced by an attenuated ERN response during easy trials and an attenuated FRN response to negative feedback during difficult trials. Further, we hypothesized that women in the SI group would demonstrate increased P300 and Pe amplitudes, indicating increased conscious attention to errors, associated with more adaptive use of performance monitoring, as compared to the ST group.

4.2 Methods

4.2.1 Participants. Fifty-seven female college students were recruited from the University of Massachusetts (UMass) Amherst. Of these participants, fourteen were excluded due to equipment failure, one was excluded because of experimenter error, and three were excluded from analysis because they were at least 3 standard deviations away from the group on key variables, leaving a final sample size of thirty-nine. Overall, excluded participants did not differ from included participants on age, $t(55) = .303$ p=.231, general anxiety, $t(54) = .788$ p=.434, general stress, $t(54) = 1.156$ p=.253, math identification score, $t(55) = -.847$ p=.401, or distribution amongst the three conditions, $X^2(3) = 2.619$, $p = 0.454$, φ=.214. Participants in Study 3 also did not differ from female participants in Study 1 or 2 in age, $F(2, 138) = 1.444$ p=.240, general stress, $F(2, 136) = 2.351$ p=.099, general anxiety, $F(2, 135) = 1.754$ p=.177, or math identification, F(2,

136) = 1.960 p=.145. To participate, students needed to be at least 18 years of age and have completed the pre-screen survey via the UMass Amherst participant recruitment SONA System. Students could not participate if they had been diagnosed with a learning or attention disability, or if they were colorblind. Students were compensated for their time with extra credit points in participating psychology courses.

4.2.2 Procedure. Participants were randomly assigned to the stereotype threat (ST) group, the stereotype inoculation (SI) group, or the control group upon arrival. Participants were brought into the lab one at a time, where a male research assistant wearing a math-related t-shirt greeted them (Stout, Dasgupta, Hunsinger, & McManus, 2011). All participants were told that they would be performing a measure of math intelligence (consistent with the ST condition from Study 1 and Study 2) as well as a test of memory to better understand how females process visual learning cues. Following consent, the participants were fitted with an EEG/ERP cap.

Participants were then asked to rate their predicted ability on the task (see Study 1 Methods). Next, based on methods used previously to induce stereotype inoculation (Stout, Dasgupta, Hunsinger, & McManus, 2011), participants read five short biographies assigned to their condition (described in more detail below). Following these readings participants completed the practice blocks for the numerical discrimination task. The timing of the task paradigm remained the same as in Study 1. After the initial practice blocks participants were offered the option of completing an additional fourth block of practice. Participants were told that this optional block of practice would help them "*better prepare for the task*", but that it was not required.

Participants then completed three test blocks of the numerical discrimination task. Again, timing of the task remained the same as described in Study 1. At the end of the three test blocks participants were asked to complete a memory check on the biography readings. Participants were told that the memory check was used to determine "*how well they remember the previous readings".* After the memory check participants were asked to rate how well they thought they performed on the task.

Following the confidence ratings, participants were informed of an optional fourth block of the task. They were told that this block was not required, but that "*past research shows that practice helps on this task*". Before deciding about whether to complete this optional block, participants were asked to rate how confident they were that they could improve their performance in the fourth block. Then participants chose whether to complete the optional fourth block of the task. After finishing the numerical discrimination task participants completed the family background demographic questionnaire (FBQ) and the abbreviated math anxiety scale (AMAS). Participants were debriefed at the end of the visit.

4.2.3 Psychophysiological recording and data reduction. The recording and data processing remained the same as described in Study 2.

4.2.4 Measures

4.2.4.1 Numerical discrimination task. The discrimination task was the same as described in Study 2, with images from four different dot ratios (1:2, 3:4, 5:6, and 7:8).

4.2.4.2 Reading. Participants read five short biographies before starting the numerical discrimination task. For the SI condition, paragraph-long biographies were created focusing on five female mathematicians using information taken from professional sources (i.e., news articles, websites). Each biography included a picture of the woman being described. Although the community recruited from was largely White, there was some racial diversity in the population (i.e. 5 percent Black, 12 percent Asian; UMasss Office of Institutional Research, 2017). Therefore, female mathematicians chosen for the biographies came from diverse racial backgrounds (i.e., White, Black, Iranian, Asian), to increase the likelihood of participants positively identifying with the women portrayed. Biographies included details of why these women entered math fields, why they chose math as a career and what they have contributed to mathematics. Biographies in the ST condition were the same, except that the names and pronouns were modified to reflect male mathematicians, and female images were replaced with male images matched on a number of characteristics. Specifically, for the male images, three potential matches were found for each female mathematician's image. These choices were then sent to eleven researchers, who voted for the male image they believed matched each female image the most on age, attractiveness, apparent kindness, apparent intelligence, and culture/place of origin. The images with the most votes were chosen as the male matches used in the ST condition. Participants in the NT condition read five paragraphs about male and female artists. The stories remained the same as in the other two conditions, but the goals and accomplishments described were changed from math- to art-based achievements. Images were a mix of those from the SI and ST conditions.

4.2.4.3 Reading/memory check. After the first three rounds of the task, and before the choice to complete the fourth round, participants were asked to complete a memory check on the previous short readings. Participants were asked for information about the biographies (i.e., names of the protagonists/innovations, what they contributed to the math field), as well as information about how much participants related to and identified with the mathematicians (i.e., which story did they relate to the most and why). This memory check served multiple functions: 1) as a refresher of the key condition-related concepts, 2) as a check that the participants paid attention to the initial readings and retained key information, and 3) as an indicator of how much the participants identified with the mathematicians.

4.2.4.4 Confidence measure. The confidence measure remained the same as described in Study 1 and 2.

4.2.4.5 Questionnaires. The questionnaires remained the same as described in Study 1 and 2. Participants completed the DASS-21, the MSLS, and the first AMAS during the SONA prescreen, and the second AMAS and the FBQ after task performance.

4.2.3 Statistical approach. The statistical approach remained largely as described in Study 2. However, as age and math anxiety did not differ between groups, they were not controlled for in the following analysis. Further, as Study 3 comprised of three groups rather than the two seen in Study 2, multiple regressions were run for each between-group linear comparison to adequately compare ST to SI, SI to control and ST to control.

4.3 Results

4.3.1 Descriptive statistics. Participants were randomly assigned to either the ST- group (male mathematician biographies), the SI group (female mathematician biographies) or the control group (mixed gender artist biographies). Of the participants included in the current analysis, 13 were assigned to the ST group, 14 were assigned to the SI group and 12 were assigned to the control group. Groups did not significantly differ in age, $F(2, 36)= .428$ $p= 0.655$, general anxiety, $F(2, 35)= .420$ $p= 0.661$, general stress, $F(2, 36)=$

.228 p= 0.798, math anxiety level, $F(2, 36)$= 1.910 p= 0.163, or math identification scores $F(2, 36)$= .456 p= 0.637 (see Table 13).

4.3.2 Task motivation/engagement. Chi-square analysis revealed no significant group differences for completing the optional fourth block of the numerical discrimination task (p=.534; see Table 14). Logistic regressions did not show a particular factor contributing to participants' decisions to complete the fourth task round.

4.3.3 Behavioral findings. One-way ANOVAs showed no overall group difference in percent correct, $F(2, 33)$= 1.072 p= 0.354, post-error slowing, $F(2, 33)$= .431 p= 0.653, errors of omission, $F(2, 34)$= 2.052 p= 0.144, task confidence, $F(2, 32)$= .927 p= 0.406, or perception of the task as a game or a test, $F(2, 36)$= .395 p= 0.677.

However, regression analysis, controlling for overall percent correct, demonstrated a group difference in the relation between math anxiety and perception of the task for the SI group compared to the ST group, b= -.188, SE= .053, β = -2.347, p=.002. A group difference was also shown for the ST group compared to the control group, b= .140, SE= .060, β = 1.799, p=.031. No group difference was found for the SI group compared to the control group, b= -.064, SE= .065, β = -.799, p=.337. Simple slopes analysis revealed a significant positive relation between math anxiety and task perception in the ST group, such that higher math anxiety predicted participants reporting that the task was more test-like, b=.125, SE= .041, β = .817, p=.012. The opposite relation was seen for participants in the SI group, with higher math anxiety predicting higher game-ratings, b= -.090, SE= .040, β = -.598, p=.046. There was no significant relation between math anxiety and task perception in the control group (p=.735).

There was also a group difference between the SI and ST groups in the relation between initial task confidence and confidence after the task, controlling for percent correct, *b*= .977, *SE*= .448, β = 2.403, *p*=.040. No group difference was found for the ST group compared to the control group, *b*= -.619, *SE*= .437, β = -1.773, *p*=.174. No group difference was found for the SI group compared to the control group, *b*= .407, *SE*= .522, β = 1.001, *p*=.445. Although neither individual group slope showed a significant relation between confidence ratings before and after the task, the SI group and the ST showed the opposite patterns. Specifically, there was a positive trend in the SI group such that higher task confidence prior to the task predicted higher task confidence after the task, *b*= .624, *SE*= .370, β = .462, *p*=.123. In contrast, the ST group showed a negative trend, so that higher initial task confidence predicted lower later task confidence, *b*= -.358, *SE*= .277, β = -.386, *p*=.226. The control group showed no relation between task confidence before and after the task (*p*=.558; see Figure 12).

4.3.4 ERN and CRN. The ERN and CRN peak values no group differences. Likewise, there was no Condition (*p*=.691) or Level x Condition (*p*=.325) effect for ERN difference scores.

Regressions showed a significant relation across all conditions, between ERN peak amplitude and percent correct, such that more ERN reactivity during easy, *b*= -1.47, *SE*= .64, β = -.37, *p*=.03, and hard trials, *b*= -2.06, *SE*= .84, β = -.39, *p*=.02, predicted higher accuracy. Similarly, more CRN reactivity during easy, *b*= -1.56, *SE*= .73, β = -.35, *p*=.04, and hard trials, *b*= -2.59, *SE*= .79, β = -.50, *p*=.003, predicted higher accuracy. No other relations were found for ERN peak amplitude or CRN peak amplitudes and performance measures (see Tables 15a and 15b respectively). Results revealed no relation between ERN difference scores and task performance (see Table 15c).

Regressions demonstrated a negative relation between ERN peak reactivity during easy trials and general stress, with higher stress predicting more attenuated ERN peaks, b= 2.29, SE= .90, β = .43, p=.02 (see Table 16a). A significant positive relation was also found across all groups between CRN reactivity during easy trials and general anxiety, such that higher anxiety predicting more attenuated CRN amplitudes, b= 2.25, SE= .94, β = .39, p=.02 (see Table 16b). Regression analyses, controlling for percent correct, revealed no relation between ERN difference scores and participant attitudes (i.e., Math ID, general anxiety, general stress scores or task confidence; see Table 16c).

4.3.5 Post-error Positivity (Pe). A 2-way interaction emerged for the peak-to-peak Pe between Trial Type (Error vs. Correct Response) x Condition (ST vs. SI vs. control), $F(2, 32) = 6.520$ $p = .004$. A condition effect was found for the Pe difference score ($Pe_{(incorrect)}$ minus $Pe_{(correct)}$), $F(2, 32) = 6.520$ $p = .004$.

Follow up ANCOVAs revealed a significant difference in reactivity to error responses compared to correct responses between the ST and the SI groups, $F(1, 23) = 5.668$ $p = .026$, and between the SI and control groups, $F(1, 20) = 9.466$ $p = .006$. There was no group difference in Pe peak-to-peak reactivity to errors versus correct responses between the ST and control groups (p=.122). Overall, this trial type x condition interaction was driven by a group difference in reactivity to correct responses, $F(2, 32) = 7.419$ $p = .002$, with the SI group showing larger Pe amplitudes (M = 2.445 SE=.266) than the ST group (M = 1.221 SE=.268) or the control group (M = 1.095 SE=.309). There was no group difference in Pe peak-to-peak reactivity to errors (p=.448; see Figure 13).

Regressions revealed a relation across all groups between $Pe_{(incorrect)}$ peak-to-peak amplitude, averaged across easy and hard trials, and errors of omission such that larger

$Pe_{(incorrect)}$ amplitudes predicted more errors of omission, *b*= 7.98, *SE*= 3.29, *β* = .38, *p*=.02. No other relations were found between $Pe_{(incorrect)}$ or $Pe_{(correct)}$ peak-to-peak scores and task performance (i.e., errors of omission, percent correct, post-error slowing; see Tables 17a and 17b respectively). Regressions revealed no relation between the Pe difference score and measures of task performance (i.e., errors of omission, percent correct, post-error slowing; see Table 17c).

$Pe_{(incorrect)}$ peak-to-peak scores during difficult errors related to math identification scores such that higher math ID predicted more attenuated $Pe_{(incorrect)}$ amplitudes, *b*= -1.886, *SE*= .771, *β* = -.389, *p*=.02. No other relations were found between participant attitudes and $Pe_{(incorrect)}$ or $Pe_{(correct)}$ peak-to-peak scores (see Tables 18a and 18b respectively). Regressions, controlling for percent correct showed no overall relations between the Pe difference score and participant attitudes (see Table 18b).

4.3.6 FRN. There was a condition effect for FRN reactivity, $F(2, 31) = 3.279 p = .05$. Follow-up ANCOVAs revealed a significant difference in FRN peak-to-peak amplitude between ST and SI conditions, $F(1, 23) = 4.471$ $p = .046$, and SI and control conditions, $F(1, 19) = 7.654$ $p = .012$, with the SI group showing more attenuated FRN peak-to-peak amplitudes overall (*M*=-2.448, *SE*=.407) compared to the ST (*M*=-3.693, *SE*=.410) and the control groups (*M*=-3.855, *SE*=.501). No group difference was evident between the ST and control conditions (*p*=.736; see Figure 14).

A significant relation was found between $FRN_{(negative)}$ peak-to-peak amplitudes during difficult trials and errors of omission, with a larger $FRN_{(negative)}$ predicting more errors of omission, *b*= -2.92, *SE*= 1.25, *β* = -37, *p*=.03 (see Table 19a). Similarly, a significant relation was found between FRN difference score and errors of omission during easy trials, b= -3.69, SE= 1.47, β = -.40, p=.017 and hard trials, b= -4.78, SE= 1.17, β = -.58, p<.001. Specifically, less FRN reactivity to negative feedback predicted fewer errors of omission (see Table 19c). No other relations between FRN peak-to-peak values or FRN difference score amplitudes and task performance were found (see Tables 19a-19c).

Regressions revealed no relation between FRN peak-to-peak amplitudes and participant attitudes (i.e., math ID, math anxiety, general anxiety, general stress scores or task confidence; see Tables 20a and 20b respectively). Similarly, no relation was found between the FRN difference score and participant attitudes (see Table 20c).

4.3.7 P300. Analysis showed no main effect of group on P300 peak-to-peak values. A significant relation was found across all three groups between $P300_{(negative)}$ amplitude and overall errors of omission during easy, b= 1.79, SE= .83, β = .34, p=.04, and hard trials, b= 2.91, SE= .71, β = .57, p=.001, such that higher P300 peak-to-peak amplitudes to errors of commission predicted more errors of omission (see Table 21a). Similar relations were found between errors of omission and $P300_{(positive)}$ peak-to-peak amplitude during difficult trials, b= 3.15, SE= 1.47, β = .34, p=.04, and P300 difference scores during easy, b= 1.93, SE= .97, β = .32, p=.05, and difficult trials b= 3.63, SE= 1.01, β = .52, p<.001 (see Tables 21b and 21c respectively).

Regression analyses revealed no relation between P300 reactivity, as measured via peak or difference score, and participant attitudes (i.e., Math ID, general anxiety, general stress scores or task confidence; see Tables 22a-22c).

4.3.8 Associations between ERN, CRN and FRN. A positive correlation between ERN peak amplitudes to easy errors and $FRN_{(negative)}$ amplitudes during easy trials was found in the SI group, r=.536 p=.048. Specifically, participants in the SI group demonstrated a positive relation between the two ERPs, such that more neural reactivity to easy errors (ERN peak) predicted more neural reactivity to negative feedback after easy errors (FRN). No further correlations were found between ERN and $\text{FRN}_{(\text{negative})}$ peaks or between CRN peaks and $\text{FRN}_{(\text{positive})}$ peaks within the SI group. No significant correlations were found for ERN and $\text{FRN}_{(\text{negative})}$ peaks or between CRN peaks and $\text{FRN}_{(\text{positive})}$ peaks within the ST or control groups.

No significant relation was found between ERN difference score amplitude and FRN difference score amplitude in any group.

4.3.9 Association between ERN, CRN and Pe. No relation was found between ERN or CRN peak amplitudes and Pe peak-to-peak amplitudes in any group.

For difference scores, a significant negative relation was found between ERN difference scores and Pe difference scores during easy trials within the ST group, r=-.570 p=.042, with more ERN difference score reactivity predicting higher Pe difference score amplitudes. No further significant relations were found between ERN difference score amplitude and Pe difference score amplitude in any group.

4.3.10 Mean score analysis. Overall ANCOVAs were run to determine whether major group findings within the peak and peak-to-peak scores remained the same when measured as mean score. The condition by trial type findings for the Pe peak-to-peak scores did not hold with mean scores, $F(2, 32) = 2.204$ $p = .127$ $\eta_p 2$=.121. Further, the condition effect for the FRN peak-to-peak scores did not hold when mean scores were used, $F(2, 31) = 1.421$ $p = .257$ $\eta_p 2$=.084.

4.4 Discussion

Study 3 aimed to explore the impact of the SIM on performance monitoring under ST using ERP methodology. The analysis, like in Study 2, focused on four ERP waveforms well linked to systems of response-monitoring (ERN/CRN and Pe) and feedback-monitoring (FRN and P300). Our results indicate that the SIM, administered in the manner described above, does not negate ST effects, but instead changes participant's perceptions and focus during the task in adaptive ways.

Behaviorally, the SIM seems to change women's perceptions toward, and relation to, the task. Women with high math anxiety reported viewing the math task as more of a game after being exposed to in-group experts. In contrast, women with high math-anxiety in the ST group reported perceiving the task as more of a test. Furthermore, women in the SI group demonstrated more consistent task confidence from the start of the task to the end, with higher initial task confidence predicting higher later task confidence. The combination of these patterns suggests that the SI condition may be acting as a buffer for women vulnerable to ST effects, allowing them to reframe their task perception to distance their self-esteem from the task outcome in a protective, and adaptive manner.

Interestingly, this pattern is similar to results seen in ST-induced task disengagement (Crocker & Major, 1989; Major, Spencer, Schmader, Wolfe & Crocker, 1998; Woodcock, Hemandez, Estrada & Schultz, 2012), however, it seems possible that this SI-induced reframing allows for the self-protective effects of ST-disengagement without the long-term risk of de-identification. ST-disengagement is believed to be a reactive measure to stereotype threat, whereas this SI-induced reframing seems to be more proactive. Indeed, as the SI reframing is linked to participants reading about successful women mathematicians who often struggled before they succeeded, these narratives may have allowed anxious individuals to separate their self-worth from the task's outcome, without compromising their overall interest and motivation in the field. This interpretation is consistent with previous behavioral effects found in studies using SI (Stout et al., 2010; Asgari, Dasgupta & Cote, 2010; Dasgupta, 2011). For example, women taught by a female calculus professor demonstrate higher confidence in their abilities and more positive attitudes towards math compared to females taught by a male teacher (Stout et al., 2010). Although Study 3 did not show a group difference in task motivation, this may have been due to the same ERP neural-net conditions described in Study 2 (i.e. slight discomfort and lengthening of lab session), which may have shifted the reasons to complete the fourth task-round from self-motivation to compliance. Therefore, further studies are needed to determine if this SI-related reframing serves to protect motivation within the threatening field.

In line with these behavioral findings and with our initial hypothesis, the SI group demonstrated a more attenuated FRN amplitude to both positive and negative feedback compared to women in the other two groups. This lower FRN reactivity suggests a general decrease in monitoring feedback among inoculated women (Forbes, Schmader & Allen, 2008; Clayson, Clawson & Larson, 2012). This lower FRN reactivity may represent less processing of external performance evaluation information, which may serve as a protective measure to buffer the effects that ST can have on self-esteem and confidence. This idea is supported by the relation between FRN amplitude and errors of omission, with more attenuated FRN predicting fewer errors of omission. This pattern suggests that women in the SI group did not attend as much to the feedback, allowing them to attend more to the task. Although our data did not demonstrate any significant relations between FRN reactivity and task confidence, this absence may be due to our small sample size. With a larger sample size, patterns between these neural measures and behavioral indicators of confidence may emerge.

Although we initially hypothesized a change in error-monitoring among women exposed to the SIM, our results indicate that these women monitor *correct* responses more than women in the ST condition. Specifically, women in the SI group show increased Pe amplitudes to both correct responses and errors while women in the other two groups show enhanced Pe amplitudes to errors alone. Typically, the Pe is larger in response to errors, representing the conscious awareness and perceived importance of an error (Overbeek, Nieuwenhuis & Ridderinkhof, 2005; Ullsperger et al., 2007). Very little is known about Pe reactivity in response to correct answers, as this phenomenon is not typically seen (Endrass, et al., 2012; Dhar, Wiersema & Pourtois, 2011). However, theories can be drawn from the neural functions of the brain regions thought to generate the Pe waveform. Specifically, source localization studies have pointed to the anterior cingulate cortex (ACC) as the source of the Pe (Herrmann, et al., 2004; Overbeek et al., 2005). Although the ACC is often associated with error processing (Brown & Braver, 2005; Magno et al., 2006), it more likely reflects the importance of information for effective learning (Silvetti, Seurinck & Verguts, 2012; Behrens et al., 2007). In many cases this process would emphasize errors, as errors are often rarer than correct answers and provide more information about how to improve task performance. However, heightened ACC activity has been shown in response to rewarding stimuli when that information is useful for overall task learning (Behrens et al., 2007; Amiez, Joseph &

Procyk, 2006). Thus, the increase in Pe amplitude to correct responses among women in the SI condition may suggest that perception of correct responses was important to this group. This increased salience of accurate responses may then be used to learn from correct responses as well as errors and improve overall task performance. This association might serve to counteract the negative ST effects on performance (Schmader et al., 2008). Although there were no relations between increased Pe to correct responses and performance measures in our data, this lack of findings may be due to our relatively small sample size. With a larger sample, patterns between Pe reactivity and performance might start to appear. This lack of relation between Pe amplitude and task performance may also have to do with the nature of the task used. As the numerical discrimination task relies on numerical approximations it requires less use of executive functioning (i.e. working memory, selective attention). Therefore, learning during the numerical discrimination task may not relate to performance monitoring to the same extent as in previous studies.

The limitations for this study remain the same as those described in Study 2. Specifically, there was an overall participant homogeneity in academic discipline and race, and a smaller sample size. Further, the peak-to-peak patterns did not remain the same when mean scores were used. However, this lack in consistency between peak-to-peak and mean findings may be due to the relation between different waveforms, which the peak-to-peak measure can uniquely elucidate. For example, the condition difference observed for the FRN peak-to-peak measures encapsulates the complete deflection from the P200 peak to the FRN peak for each group. Interestingly, it is the relation between these two waveforms that seems to differ by group, rather than the FRN in isolation.

Specifically, while both the ST and control groups show distinct deflections from the P200 peak to the FRN peak, the SI group shows only a slight change in amplitude between these two waveforms (see Figure 14). This relation cannot be accurately measured by looking at the FRN alone, which is what the mean amplitude is measuring. Therefore, in this study, the peak-to-peak measure seems to be more accurately showing group differences in reactivity than the mean measures.

Overall, this study offers an important first step towards understanding *how* SI protects women under ST. Importantly, these findings suggest that the mechanisms through which SI work are not a simple reversal of ST effects. Instead, SI may function to shift women's perceptions of threatening tasks, lowering their responses to external feedback and increasing their focus on internal monitoring of correct answers. Further studies are still needed to determine how differing methods of implementing SI (i.e. in-person role-models, more extended exposure to role-models, etc.) may differentially impact ST effects.

Table 13a. Means and standard errors for age, anxiety scores, stress scores, math anxiety level and math identification scores in Study 3

	Stereotype Inoculation	Stereotype Threat	Control
Final Sample Size	14	13	12
Age: Mean *(SE)*	19.57*(0.45)*	19.69*(0.23)*	19.25*(0.25)*
Anxiety: Mean *(SE)*	10.86*(2.32)*	9.83*(2.42)*	13.17*(2.99)*
Stress: Mean *(SE)*	14.57*(2.15)*	12.62*(2.55)*	15.00*(3.28)*
Math Anxiety Level: Mean *(SE)*	1.71*(0.19)*	2.23*(0.20)*	2.08*(0.19)*
Math ID: Mean *(SE)*	59.07*(1.49)*	4.75*(1.32)*	57.67*(0.86)*

Table 13b. Means, standard errors and between group comparisons (p-values) for the number of epochs acquired following response onset and feedback onset, controlling for reaction time and time of day in Study 3.

Trigger Type	Trial Level	Stereotype Inoculation Mean*(SE)*	Stereotype Threat Mean*(SE)*	Control Mean*(SE)*	*p*
Error Response	Easy	43.00*(21.14)*	42.69*(20.22)*	54.00*(23.59)*	.580
	Hard	78.69*(18.63)*	74.46*(21.00)*	86.40*(13.89)*	.640
Correct Answer Response	Easy	183.46*(25.63)*	190.08*(47.83)*	181.80*(53.38)*	.996
	Hard	142.15*(20.31)*	152.15*(39.31)*	148.80*(36.37)*	.857
Negative Feedback	Easy	42.23*(20.93)*	40.06*(15.90)*	54.10*(23.42)*	.472
	Hard	77.46*(17.96)*	72.83*(20.89)*	85.30*(14.72)*	.571
Positive Feedback	Easy	180.92*(26.33)*	187.77*(49.72)*	178.40*(52.95)*	.994
	Hard	141.23*(20.35)*	149.17*(38.80)*	148.20*(36.16)*	.887

Table 14. Number of participants who chose to complete the optional fourth task round per group in Study 3.

	Stereotype Inoculation	Stereotype Threat	Control
Yes to 4^{th} Round	4	5	6
No to 4^{th} Round	10	8	6

Table 15a. Regressions between task performance measures and ERN reactivity to easy and hard trials in Study 3.

Performance Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Errors of Omission					
	Easy	-0.87 *(1.78)*	-.08	-0.49	.63
	Hard	0.35 *(2.35)*	.03	0.15	.88
Percent Correct					
	Easy	-1.47 *(0.64)*	-.37	-2.30	.03*
	Hard	-2.06 *(0.84)*	-.39	-2.46	.02*
Post-Error Slowing					
	Easy	-6.33 *(3.46)*	-.30	-1.83	.08
	Hard	-8.64 *(4.56)*	-.31	-1.90	.07

*$p<.05$

Table 15b. Regressions between task performance measures and CRN reactivity to easy and hard trials in Study 3.

Performance Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Errors of Omission					
	Easy	-2.10 *(1.95)*	-.18	-1.08	.29
	Hard	-2.74 *(2.34)*	-.20	-1.17	.25
Percent Correct					
	Easy	-1.56 *(0.73)*	-.35	-2.15	.04*
	Hard	-2.59 *(0.79)*	-.50	-3.27	.003*
Post-Error Slowing					
	Easy	-4.57 *(4.02)*	-.19	-1.14	.26
	Hard	-7.24 *(4.64)*	-.26	-1.56	.13

*$p<.05$

Table 15c. Regressions between task performance measures and ERN difference scores (ERN minus CRN) to easy and hard trials in Study 3.

Performance Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Errors of Omission					
	Easy	1.08 *(2.02)*	.09	0.54	.60
	Hard	6.49 *(3.27)*	.32	1.99	.06
Percent Correct					
	Easy	-0.36 *(0.84)*	-.08	-0.43	.67
	Hard	1.19 *(1.35)*	.15	0.88	.39
Post-Error Slowing					
	Easy	-4.08 *(4.37)*	-.16	-0.93	.36
	Hard	-3.21 *(7.21)*	-.08	-0.45	.66

*$p<.05$

Table 16a. Regressions between participant attitude measures and ERN reactivity to easy and hard trials, controlling for percent correct, in Study 3.

Attitude Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Math ID					
	Easy	-0.61 *(0.50)*	-.23	-1.24	.23
	Hard	-0.33 *(0.68)*	-.09	-0.50	.62
General Anxiety					
	Easy	1.57 *(0.86)*	.31	1.83	.08
	Hard	1.53 *(1.22)*	.22	1.26	.22
General Stress					
	Easy	2.29 *(0.90)*	.43	2.55	.02*
	Hard	1.41 *(1.29)*	.20	1.09	.28
Task Confidence					
	Easy	-0.05 *(0.04)*	-.24	-1.27	.21
	Hard	-0.05 *(0.05)*	-.18	-0.98	.34

*$p<.05$

Table 16b. Regressions between participant attitude measures and CRN reactivity to easy and hard trials, controlling for percent correct, in Study 3.

Attitude Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Math ID					
	Easy	-0.21 *(0.57)*	-.07	-0.37	.71
	Hard	0.21 *(0.72)*	.06	0.29	.77
General Anxiety					
	Easy	2.25 *(0.94)*	.39	2.40	.02*
	Hard	2.19 *(1.24)*	.32	1.77	.09
General Stress					
	Easy	1.87 *(1.05)*	.31	1.78	.09
	Hard	2.11 *(1.34)*	.30	1.58	.13
Task Confidence					
	Easy	-0.04 *(0.05)*	-.15	-0.81	.42
	Hard	-0.02 *(0.06)*	-.08	-0.40	.69

$^*p<.05$

Table 16c. Regressions between participant attitude measures and ERN difference scores (ERN minus CRN) to easy and hard trials, controlling for percent correct, in Study 3.

Attitude Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Math ID					
	Easy	-0.58 *(0.57)*	-.18	-1.01	.32
	Hard	-1.02 *(0.93)*	-.19	1.10	.28
General Anxiety					
	Easy	-0.25 *(1.03)*	-.04	-0.24	.81
	Hard	-0.90 *(1.69)*	-.09	-0.53	.60
General Stress					
	Easy	1.03 *(1.11)*	.16	0.94	.36
	Hard	-0.93 *(1.84)*	-.09	-0.51	.62
Task Confidence					
	Easy	-0.03 *(0.05)*	-.11	-0.61	.55
	Hard	-0.07 *(0.08)*	-.15	-0.84	.41

*$p<.05$

Table 17a. Regressions between task performance measures and $Pe_{(incorrect)}$ reactivity to easy and hard trials in Study 3.

Performance Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Errors of Omission					
	Easy	4.36 *(2.44)*	.29	1.79	.08
	Hard	6.25 *(3.06)*	.33	2.04	.05*
Percent Correct					
	Easy	0.91 *(1.00)*	.15	0.91	.37
	Hard	-0.41 *(1.22)*	-.06	-0.34	.74
Post-Error Slowing					
	Easy	3.64 *(5.33)*	.12	0.68	.50
	Hard	-3.08 *(6.42)*	-.08	-0.48	.63

*$p<.05$

Table 17b. Regressions between task performance measures and $Pe_{(correct)}$ reactivity to easy and hard trials in Study 3.

Performance Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Errors of Omission					
	Easy	4.87 *(2.72)*	.29	1.79	.08
	Hard	3.67 *(2.52)*	.24	1.46	.15
Percent Correct					
	Easy	-0.56 *(1.07)*	-.09	-0.52	.61
	Hard	-0.82 *(0.97)*	-.14	-0.85	.40
Post-Error Slowing					
	Easy	2.15 *(5.69)*	.07	0.38	.71
	Hard	0.73 *(5.16)*	.02	0.14	.89

$^{*}p<.05$

Table 17c. Regressions between task performance measures and Pe difference scores ($Pe_{(incorrect)}$ minus $Pe_{(correct)}$) to easy and hard trials in Study 3.

Performance Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Errors of Omission					
	Easy	0.35 *(2.32)*	.03	0.15	.88
	Hard	0.58 *(3.44)*	.03	0.17	.87
Percent Correct					
	Easy	1.08 *(0.87)*	.21	1.24	.22
	Hard	0.98 *(1.28)*	.13	0.76	.45
Post-Error Slowing					
	Easy	1.33 *(4.70)*	.05	0.28	.78
	Hard	-4.76 *(6.80)*	-.12	-0.70	.49

*$p<.05$

Table 18a. Regressions between participant attitude measures and $Pe_{(incorrect)}$ reactivity to easy and hard trials, controlling for percent correct, in Study 3.

Attitude Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Math ID					
	Easy	-0.36 *(0.70)*	-.09	-0.51	.62
	Hard	-1.89 *(0.77)*	-.39	-2.45	.02*
General Anxiety					
	Easy	-0.71*(1.25)*	-.09	-0.57	.57
	Hard	1.35 *(1.51)*	.14	0.89	.38
General Stress					
	Easy	-2.15 *(1.34)*	-.27	-1.61	.12
	Hard	1.56 *(1.63)*	.16	0.96	.35
Task Confidence					
	Easy	-0.09 *(0.06)*	-.28	-1.64	.11
	Hard	-0.06 *(0.07)*	-.15	-0.85	.41

*$p<.05$

Table 18b. Regressions between participant attitude measures and $Pe_{(correct)}$ reactivity to easy and hard trials, controlling for percent correct, in Study 3.

Attitude Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Math ID					
	Easy	-0.85 *(0.73)*	-.20	-1.16	.25
	Hard	-1.08 *(0.65)*	-.28	-1.67	.11
General Anxiety					
	Easy	0.28 *(1.33)*	.04	0.21	.83
	Hard	-0.39 *(1.21)*	-.05	-0.32	.75
General Stress					
	Easy	0.50 *(1.46)*	.06	0.34	.73
	Hard	-0.51 *(1.33)*	-.07	-0.38	.71
Task Confidence					
	Easy	-0.03 *(0.06)*	-.10	-0.57	.57
	Hard	-0.03 *(0.05)*	-.10	-0.56	.58

*$p<.05$

Table 18c. Regressions between participant attitude measures and Pe difference scores ($Pe_{(incorrect)}$ minus $Pe_{(correct)}$) to easy and hard trials, controlling for percent correct, in Study 3.

Attitude Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Math ID					
	Easy	0.32 *(0.62)*	.09	0.51	.61
	Hard	-0.26 *(0.90)*	-.05	-0.29	.77
General Anxiety					
	Easy	-0.76 *(1.11)*	-.11	-0.69	.50
	Hard	2.18 *(1.57)*	.22	1.39	.17
General Stress					
	Easy	-2.05 *(1.18)*	-.29	-1.74	.09
	Hard	2.67 *(1.71)*	.26	1.57	.13
Task Confidence					
	Easy	-0.05 *(0.05)*	-.16	-0.91	.37
	Hard	-0.01 *(0.07)*	-.03	-0.15	.88

*$p<.05$

Table 19a. Regressions between task performance measures and $FRN_{(negative)}$ reactivity to easy and hard trials in Study 3.

Performance Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Errors of Omission					
	Easy	-2.32 *(1.31)*	-.29	-1.77	.09
	Hard	-2.92 *(1.25)*	-.37	-2.34	.03*
Percent Correct					
	Easy	-0.57 *(0.51)*	-.19	-1.11	.28
	Hard	-0.69 *(0.50)*	-.23	-1.37	.18
Post-Error Slowing					
	Easy	-0.02 *(2.76)*	-.001	-0.01	.99
	Hard	-0.51 *(2.72)*	-.03	-0.19	.85

*$p<.05$

Table 19b. Regressions between task performance measures and $FRN_{(positive)}$ reactivity to easy and hard trials in Study 3.

Performance Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Errors of Omission					
	Easy	0.80 *(1.97)*	.07	0.40	.69
	Hard	4.76 *(2.39)*	.32	1.98	.06
Percent Correct					
	Easy	-0.52 *(0.75)*	-.12	-0.70	.49
	Hard	0.53 *(0.97)*	.10	0.55	.59
Post-Error Slowing					
	Easy	0.69 *(3.99)*	.03	0.17	.86
	Hard	0.20 *(5.13)*	.01	0.04	.97

*$p<.05$

Table 19c. Regressions between task performance measures and FRN difference scores ($FRN_{(negative)}$ minus $FRN_{(positive)}$) to easy and hard trials in Study 3.

Performance Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Errors of Omission					
	Easy	-3.69 *(1.47)*	-.40	-2.51	.02*
	Hard	-4.78 *(1.17)*	-.58	-4.10	<.001*
Percent Correct					
	Easy	-0.43 *(0.61)*	-.12	-0.71	.48
	Hard	-0.95 *(0.52)*	.30	-1.81	.08
Post-Error Slowing					
	Easy	-0.48 *(3.22)*	-.03	-0.15	.88
	Hard	-0.65 *(2.89)*	-.04	-0.22	.83

*$p<.05$

Table 20a. Regressions between participant attitude measures and $FRN_{(negative)}$ reactivity to easy and hard trials, controlling for percent correct, in Study 3.

Attitude Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Math ID					
	Easy	-0.53 *(0.41)*	-.23	-1.31	.20
	Hard	-0.52 *(0.38)*	-.24	-1.34	.19
General Anxiety					
	Easy	-0.27 *(0.76)*	-.06	-0.34	.72
	Hard	-0.46 *(0.72)*	-.11	-0.64	.53
General Stress					
	Easy	0.57 *(0.81)*	.12	0.71	.49
	Hard	0.50 *(0.76)*	.12	0.66	.51
Task Confidence					
	Easy	-0.01 *(0.03)*	-.07	-0.38	.71
	Hard	0.01 *(0.03)*	.05	0.28	.78

*$p<.05$

Table 20b. Regressions between participant attitude measures and $FRN_{(positive)}$ reactivity to easy and hard trials, controlling for percent correct, in Study 3.

Attitude Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Math ID					
	Easy	-0.53 *(0.41)*	-.23	-1.31	.20
	Hard	-0.52 *(0.38)*	-.24	-1.34	.19
General Anxiety					
	Easy	-0.27 *(0.76)*	-.06	-0.34	.72
	Hard	-0.46 *(0.72)*	-.11	-0.64	.53
General Stress					
	Easy	0.57 *(0.81)*	.12	0.71	.49
	Hard	0.50 *(0.76)*	.12	0.66	.51
Task Confidence					
	Easy	-0.01 *(0.03)*	-.07	-0.38	.71
	Hard	0.01 *(0.03)*	.05	0.28	.78

*$p<.05$

Table 20c. Regressions between participant attitude measures and FRN difference scores ($FRN_{(negative)}$ minus $FRN_{(positive)}$) to easy and hard trials, controlling for percent correct, in Study 3.

Attitude Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Math ID					
	Easy	-0.53 *(0.41)*	-.23	-1.31	.20
	Hard	-0.52 *(0.38)*	-.24	-1.34	.19
General Anxiety					
	Easy	-0.27 *(0.76)*	-.06	-0.34	.72
	Hard	-0.46 *(0.72)*	-.11	-0.64	.53
General Stress					
	Easy	0.57 *(0.81)*	.12	0.71	.49
	Hard	0.50 *(0.76)*	.12	0.66	.51
Task Confidence					
	Easy	-0.01 *(0.03)*	-.07	-0.38	.71
	Hard	0.01 *(0.03)*	.05	0.28	.78

*$p<.05$

Table 21a. Regressions between task performance measures and $P300_{(negative)}$ reactivity to easy and hard trials in Study 3.

Performance Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Errors of Omission					
	Easy	1.79 *(0.83)*	.34	2.16	.04*
	Hard	2.91 *(0.71)*	.57	4.08	<.001*
Percent Correct					
	Easy	0.50 *(0.32)*	.26	1.55	.13
	Hard	0.78 *(0.30)*	.41	2.62	.01
Post-Error Slowing					
	Easy	-0.67 *(1.76)*	-.07	-0.38	.71
	Hard	0.11 *(1.73)*	.01	0.06	.95

*$p<.05$

Table 21b. Regressions between task performance measures and $P300_{(positive)}$ reactivity to easy and hard trials in Study 3.

Performance Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Errors of Omission					
	Easy	1.33 *(1.71)*	.13	0.77	.44
	Hard	3.15 *(1.47)*	.34	2.14	.04*
Percent Correct					
	Easy	0.74 *(0.64)*	.19	1.15	.26
	Hard	1.11 *(0.56)*	.32	1.99	.06
Post-Error Slowing					
	Easy	1.97 *(3.44)*	.10	0.57	.57
	Hard	3.07 *(3.07)*	.17	1.00	.32

*$p<.05$

Table 21c. Regressions between task performance measures and P300 difference scores ($P300_{(negative)}$ minus $P300_{(positive)}$) to easy and hard trials in Study 3.

Performance Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Errors of Omission					
	Easy	1.93 *(0.97)*	.32	1.99	.05*
	Hard	3.63 *(1.01)*	.52	3.58	<.001*
Percent Correct					
	Easy	0.41 *(0.38)*	.18	1.09	.28
	Hard	0.83 *(0.43)*	.32	1.95	.06
Post-Error Slowing					
	Easy	-1.57 *(2.01)*	-.13	-0.78	.44
	Hard	-1.58 *(2.36)*	-.11	-0.67	.51

*$p<.05$

Table 22a. Regressions between participant attitude measures and $P300_{(negative)}$ reactivity to easy and hard trials, controlling for percent correct, in Study 3.

Attitude Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Math ID					
	Easy	0.28 *(0.21)*	.22	1.37	.18
	Hard	0.24 *(0.20)*	.19	1.20	.24
General Anxiety					
	Easy	-0.80 *(0.40)*	-.32	-2.01	.05
	Hard	-0.60 *(0.39)*	-.25	-1.54	.13
General Stress					
	Easy	-0.77 *(0.41)*	-.29	-1.87	.07
	Hard	-0.54 *(0.41)*	-.21	-1.33	.19
Task Confidence					
	Easy	0.004 *(0.02)*	.04	0.21	.84
	Hard	-0.03 *(0.02)*	-.32	-1.93	.06

$*p<.05$

Table 22b. Regressions between participant attitude measures and $P300_{(positive)}$ reactivity to easy and hard trials, controlling for percent correct, in Study 3.

Attitude Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Math ID					
	Easy	0.19 *(0.43)*	.07	0.44	.67
	Hard	0.23 *(0.38)*	.10	0.61	.54
General Anxiety					
	Easy	-0.53 *(0.84)*	-.11	-0.63	.53
	Hard	-0.56 *(0.76)*	-.12	-0.73	.47
General Stress					
	Easy	-0.77 *(0.86)*	-.15	-0.90	.38
	Hard	-0.77 *(0.77)*	-.16	-0.99	.33
Task Confidence					
	Easy	-0.01 *(0.04)*	-.03	-0.16	.88
	Hard	-0.02 *(0.03)*	-.12	-0.72	.48

*$p<.05$

Table 22c. Regressions between participant attitude measures and P300 difference scores ($P300_{(negative)}$ minus $P300_{(positive)}$) to easy and hard trials, controlling for percent correct, in Study 3.

Attitude Measure	Trial Level	B*(SE)*	ß	*t*	*p*
Math ID					
	Easy	0.29 *(0.26)*	.19	1.12	.27
	Hard	0.32 *(0.32)*	.18	1.00	.32
General Anxiety					
	Easy	-0.49 *(0.47)*	-.17	-1.03	.31
	Hard	-0.12 *(0.58)*	-.04	-0.21	.83
General Stress					
	Easy	-0.42 *(0.52)*	-.14	-0.79	.43
	Hard	-0.02 *(0.64)*	-.01	-0.03	.97
Task Confidence					
	Easy	0.01 *(0.02)*	.14	0.79	.44
	Hard	-0.04 *(0.03)*	-.26	-1.47	.15

*$p<.05$

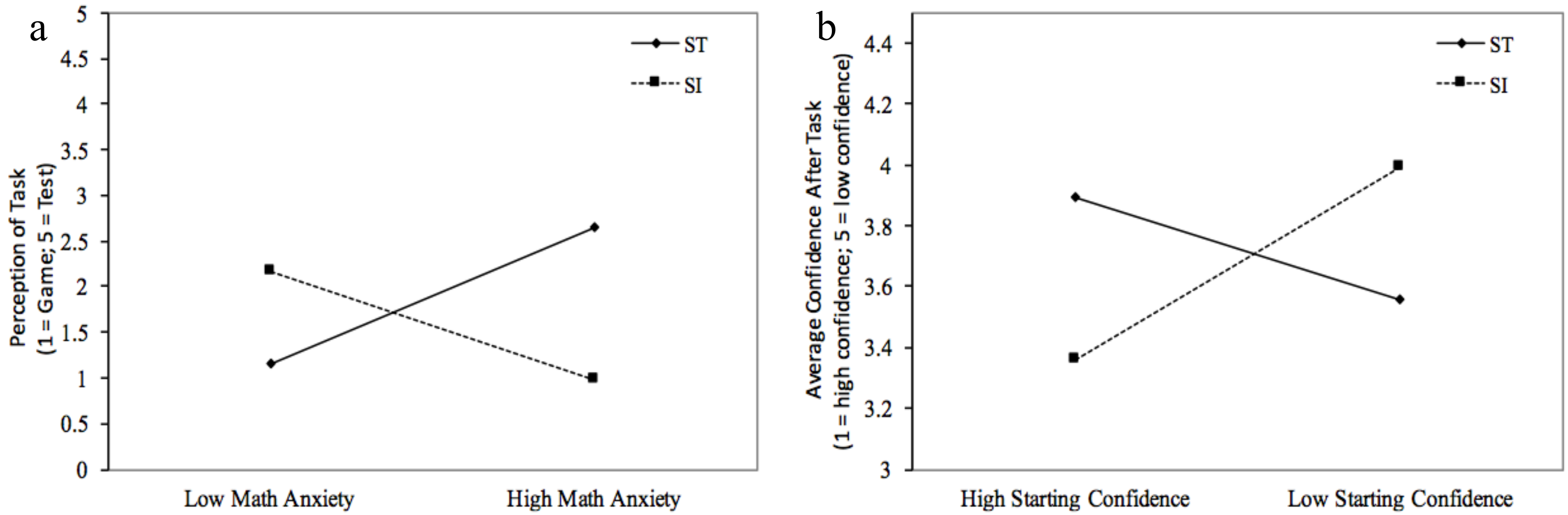


Figure 12. Predicted group (ST vs. SI) differences in the relation between a) math anxiety and perception of the task and b) starting confidence and confidence after the task. Percent correct was controlled for in these analyses.

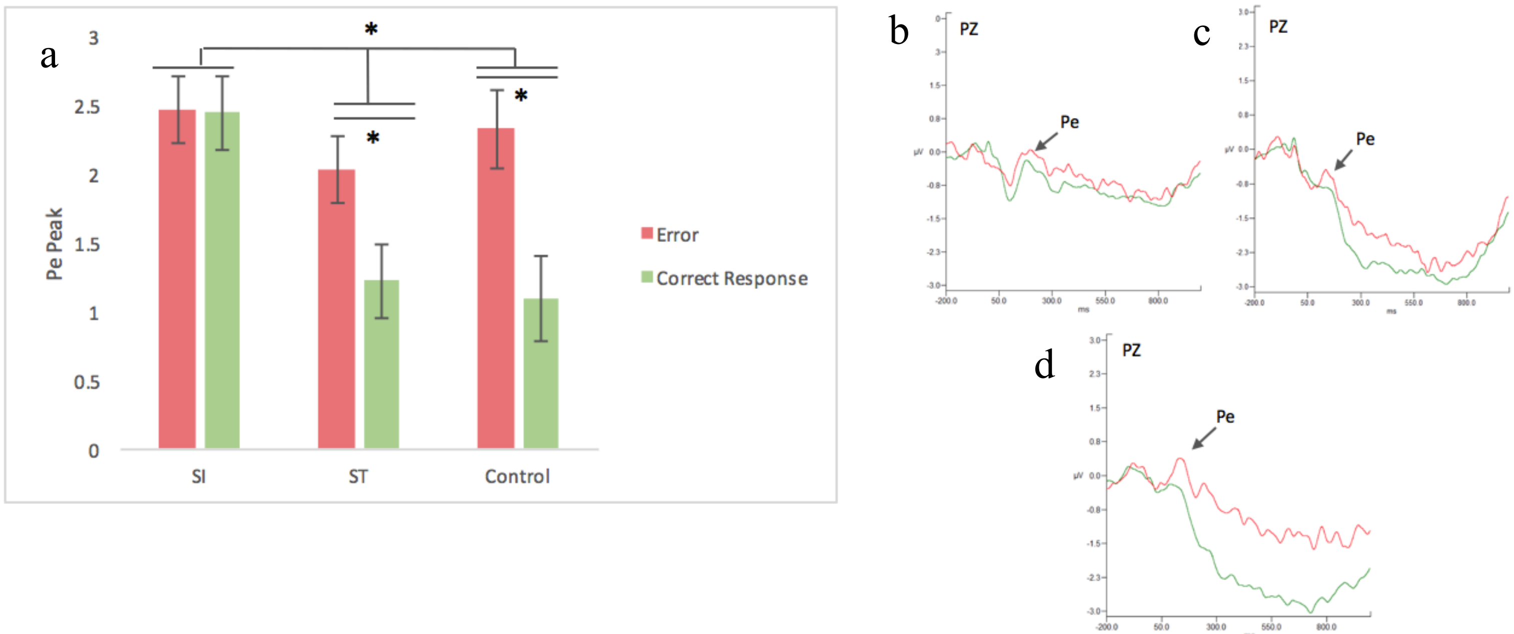


Figure 13. Pe peaks at the central-parietal region (PZ, P1, P2). a) Pe peak values, averaged across sites PZ, P1, P2, controlling for percent correct. Green bars represent Pe peaks to correct responses; red bars represent Pe peaks to errors; error bars represent standard errors; stars (*) represent significant differences. b) Average Pe waves in the SI group at site PZ. Red waves are in response to errors; green waves are in response to correct responses. Pe values used for analysis were scored as the most positive point from 50-250ms post response. c) Average Pe waves in the ST group at site PZ. d) Average Pe waves in the control group at site PZ.

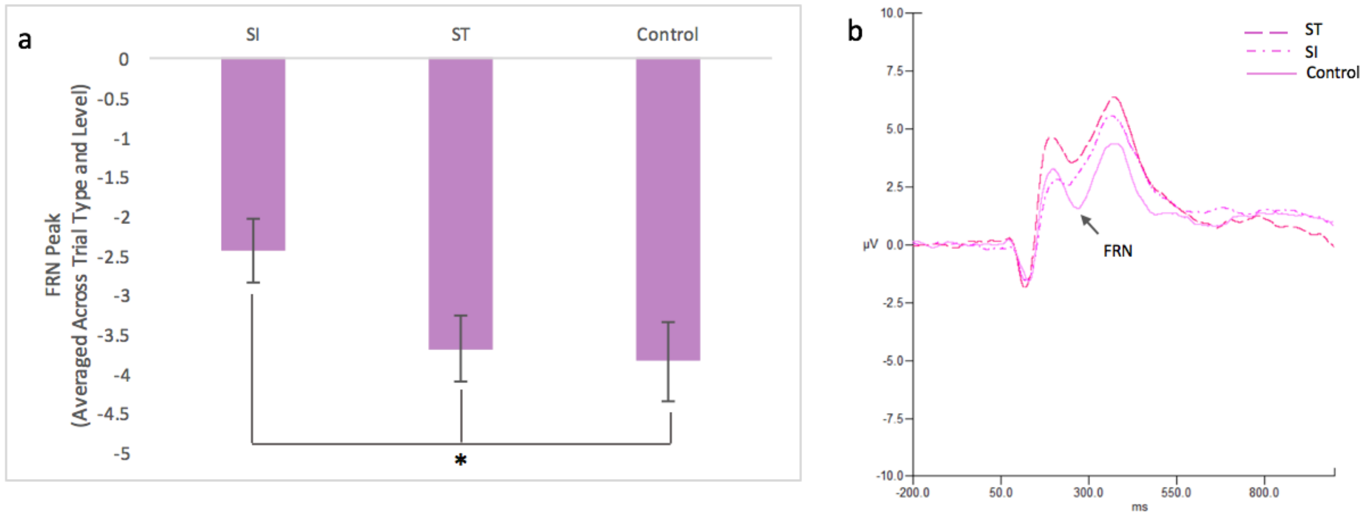


Figure 14. FRN peaks at the frontal-central region (FCZ, FC1 and FC2), averaging across level (easy and hard) and trial type (negative and positive feedback). a) Averaged FRN peak values, averaged across sites FC1, FCZ and FC2, controlling for percent correct. Error bars represent standard errors. Stars (*) represent significant differences. b) Average response waves in the SI group (dash-dot line), the ST group (dash line) and control group (solid line) at site FCZ. FRN values used for analysis were scored as the most negative point from 200-375ms post feedback.

CHAPTER 5

OVERALL DISCUSSION AND FUTURE DIRECTIONS

5.1 ST effects during a novel task

The first two studies examined the impact of ST during a novel task at the behavioral (Study 1) and neural (Study 2) levels. Women in the ST condition were less motivated to continue the task and demonstrated lower confidence in their task ability, compared to women in the NT condition and men in either condition (Study 1). Further, women in the ST group demonstrated worse task performance with higher reported anxiety and stress scores, while women in the NT group showed no such relation (Study 2). These findings are consistent with previous ST literature (Schmader, Johns & Forbes, 2008; Johns, Inzlicht & Schmader, 2010). However, previous studies have focused on ST in inherently threatening tasks, such as GRE-type math tests. These more common math tests can be influenced by previously learned helplessness linked to specific math-related symbols and commonly known tasks (Mangels, Good, Whitman, Maniscalco & Dweck, 2012), impacting the ST effect. Thus, Studies 1 and 2 add to this ST literature by showing that even during a novel task, separate from any prior practice effects, learned helplessness or experience, women's motivation, confidence and performance are negatively impacted by ST.

However, Study 1 also demonstrated that, under typical behavioral testing conditions, math identification may act to protect women's task motivation and confidence from the negative influence of ST. Specifically, women who reported higher math identification prior to the study session demonstrated a greater willingness to complete an extra task round, and showed heightened task confidence under ST conditions. These findings support previous literature demonstrating that domain identification is positively related to intrinsic motivation (Walker, Greene & Mansell, 2006), and suggest that math identification may help protect women's drive against ST effects. Further studies are needed to understand when domain identification can act as a protector against ST and when it is a vulnerability. By better understanding the full impact of domain identification we can determine how to implement domain identification in intervention work.

Beyond the behavioral findings, important patterns were revealed in the ERP waveforms (Study 2). Women under stereotype threat demonstrated enhanced, but seemingly inefficient, internal error-monitoring during easier trials, with larger ERN and CRN amplitudes. However, women in the ST condition also demonstrated more attenuated ERN amplitudes during difficult trials in relation to heightened math identification, anxiety and stress. The pattern of initial error processing suggests that stereotype threat hinders cognitive efficiency in early performance monitoring during easy trials, similar to the ERN and CRN patterns seen in anxious individuals (Endrass, Klawohn, Schuster, & Kathmann, 2008; Tanovic, Hajcak & Sanislow, 2017). However, during difficult trials, women under ST who identify with the math field and report higher anxiety and stress reactivity, may be hindered by the pressure to show their capabilities, and may subsequently have fewer mental resources to devote to task-related abilities such as performance monitoring. This ERP pattern for difficult trials supports and expands on previous behavioral studies demonstrating that women who identify with a threatening field have difficulty performing on harder ST-related task questions (Aronson, Lustina, Good, Keough, Steele & Brown, 1999; Keller, 2007).

In addition to the ERN and CRN findings, important patterns were also demonstrated for the Pe component. Namely, under ST conditions women demonstrated heightened Pe amplitudes to errors during easy trials. This enhanced Pe to errors among the ST participants then predicted lower task accuracy and more errors of omission during the task, suggesting either an inability to process the full cognitive load or more task disengagement. If these results are due to an overwhelming cognitive load, it would suggest an exhaustion of mental resources, leading to poor task performance. In contrast, disengagement would be a more conscious choice to distance from the task to protect cognitive resources and self-esteem. Future studies will explore which mechanism links the increased Pe and poor task performance under ST. For example, participants could be asked to complete an additional task, separate from the ST paradigm. If the enhanced Pe is linked to an overwhelming cognitive load, performance should suffer even on separate tasks. However, if the enhanced error-monitoring under ST induces task disengagement, then performance should be preserved on non-ST-related tasks.

However, while the findings from Studies 1 and 2 support and expand on previous ST literature, several findings from Study 1 were not replicated in Study 2. For instance, in Study 1 women under ST were less likely to complete an optional round of the task. However, in Study 2 women in both conditions were equally unlikely to complete the optional task round. The change in these results may have been impacted by necessary changes in the protocol with the addition of the ERP process. For example, the ratio of female participants to male researchers changed from Study 1 to Study 2. In Study 1, only one researcher interacted with one to four participants at a time, whereas in Study 2, two researchers were required for the ERP capping process, and only one participant could be run through the experiment at a time. Previous studies have demonstrated that the group composition can impact the perception of ST effects (Sekaquaptewa & Thompson, 2003; Murphy, Steele & Gross, 2007; Dasgupta, Scicle & Hunsinger, 2015). Specifically, when women are outnumbered by men ST effects are heightened. Therefore, the change in group composition from Study 1 to Study 2 may have impacted our paradigm by highlighting ST effects in both the ST and NT groups.

Furthermore, the addition of the ERP cap may have acted as a stressor by emphasizing the 'experiment' context and the field of science broadly, thus undermining the effectiveness of the stereotype-neutral condition. Previous literature demonstrates that situational cues, including objects in the experimental room, can heighten ST effects (Stone & McWhinnie, 2008; Murphy, Steele & Gross, 2007; Cheryan et al., 2009). The changes in the study paradigm in Study 2 may have made the NT condition more threatening than in Study 1, contributing to the lack of replication of behavioral findings in Study 2.

The increase in possible stress-inducing cues in Study 2 may also explain the change in the relation between participant attitude measures and the decision to complete the fourth task round. In Study 1 participants were more likely to complete the fourth round if they were performing well (NT group), or if they identified with the task domain (ST group). In contrast, in Study 2, participant were more likely to complete the fourth round if they reported lower confidence in their task ability. Lower self-confidence predicts more compliance, particularly among women (Gudjonsson et al., 2002; Gudjonsson & Sigurdsson, 2003). Thus, the addition of potentially stressful cues in Study 2 may have decreased the participant's self-motivation to complete the fourth round, and instead increased the relation between low self-confidence and compliance with the researchers. The ERP capping protocol in ST paradigms may also help explain why previous ERP studies exploring ST have shown inconsistent results (Forbes & Leitner, 2014; Mangels et al., 2008; Forbes, Schmader & Allen, 2008).

As no prior literature that we know of has explored the change in ST impact from a purely behavioral paradigm to an ERP paradigm, these changes in results from Study 1 to Study 2 demonstrate a new and important factor to consider in future studies exploring the neural mechanisms of ST. Future studies should explore which specific factors in the capping paradigm most influence the ST effects. For instance, the novelty and perceived intimidation of the capping procedure may impact ST effects. Thus, it may be beneficial to recruit only participants who have participated in ERP projects before, to lessen the possible impact of the capping procedure's novelty. Also, the ratio of researchers to participants may impact participant's comfort levels within the study paradigm. Future studies should explore the use of confederates to make the ratio of participant-to-researcher appear more evenly divided in ERP paradigms. Overall, further studies should aim to minimize the impact of the capping paradigm and context to more consistently represent the mechanisms underlying typical ST effects.

5.2 Insights from the SI

The final study explored how SI may impact the behavioral and neural effects of ST by adding a SI manipulation to the procedure for the ST condition from Studies 1 and 2. The SI used in Study 3 impacted both behavioral and neural patterns in ways that may indicate a SI-induced distancing of women's sense of self-worth from task outcome. Behaviorally, the SI allowed more women higher in math anxiety to reframe the task as a game, rather than a test, which may serve to lessen the impact of their task performance on their confidence and self-esteem. Similarly, the ERP patterns indicate that women exposed to examples of successful female mathematicians are more aware of correct answers during internal performance monitoring processes (enhanced Pe to correct responses), without losing their attention to saliency of their errors (equally high Pe amplitude to incorrect responses).

Like the results seen in Study 2, women in Study 3 showed more errors of omission with higher Pe amplitudes to errors, suggesting either disengagement from the task or increased cognitive load with heightened error sensitivity. Within the SI group, the Pe pattern of increased amplitudes to correct responses in addition to the reactivity to errors, indicates a possible protective mechanism, such that women exposed to the SI perceived their correct responses as equally important to their errors. Furthermore, women exposed to the SI attended less to external feedback (attenuated FRN). In related anxieties, such as social phobia, anxious individuals show a negative-bias towards evaluative feedback, including an overemphasis of negative feedback directed at the self (Brozovich & Heimberg, 2008; Morgan & Banerjee, 2008; Cody & Teachman, 2010). Thus, this attenuated feedback processing in women exposed to the SI may indicate a protective measure to lessen the impact of negative feedback bias under stress, and buffer ST effects on self-esteem and confidence. Importantly, the current findings suggest that the mechanisms through which this version of the SIM function are not a simple reversal of stereotype threat effects, but rather function to shift women's perceptions of threatening tasks, lowering their responses to external feedback and increasing their focus on internal monitoring of correct responses.

Future studies should explore the differing mechanisms for variations in SI. Previous studies have shown that exposing women to in-group role models in different ways can change the behavioral effects of the SI. For example, exposing women to role-models for a prolonged period (i.e., over the course of a semester) has more benefit for women's self-confidence compared to a one-time exposure (Stout et al., 2010). Future studies should explore the mechanistic impact of more prolonged, consistent exposure to in-group role-models. It is possible that a one-time exposure shifts task perception as in Study 3, whereas more prolonged exposure may work to counter-act the changes in error-processing observed under ST. As ST is built on years of subtle societal cues, countering it and providing a more lasting increase in confidence may require more consistent exposure to SI manipulations that build upon each other over time. Understanding this differentiation would then help policy makers and educators better implement the SIM and similar intervention techniques to protect women's math and science identification and improve their motivation to continue within the STEM fields.

5.3 Conclusion

The three studies explored in this dissertation thesis provide important insight regarding behavioral and neural mechanisms involved in ST and SI. These studies also highlight future routes of exploration needed to further elucidate mechanisms underlying ST (i.e., limiting the impact of the ERP cap to get a more natural measure of ST mechanisms), and how intervention approaches buffer detrimental effects of ST (i.e., different types of SI and over different lengths of time). Overall, these findings contribute to the understanding of the negative impact that gender-based stereotypes have on women's motivation, confidence and performance during STEM-related tasks, and the possible protective factors of math identification and exposure to in-group role models. More importantly, our results underscore the need for more evidence-based approaches to address gender disparities in STEM-related fields.

REFERENCES

Ambady, N., Paik, S. K., Steele, J., Owen-Smith, A., & Mitchell, J. P. (2004). Deflecting negative self-relevant stereotype activation: The effects of individuation. *Journal of Experimental Social Psychology*, *40*(3), 401-408.

Amodio, D. M. (2010). Can neuroscience advance social psychological theory? Social neuroscience for the behavioral social psychologist. *Social Cognition*, *28*(6), 695.

Amiez, C., Joseph, J. P., & Procyk, E. (2005). Reward encoding in the monkey anterior cingulate cortex. *Cerebral cortex*, *16*(7), 1040-1055.

Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of experimental social psychology*, *35*(1), 29-46.

Aronson, J., Quinn, D. M., & Spencer, S. J. (1998). *Stereotype threat and the academic underperformance of minorities and women.* San Diego, CA: Academic Press

Asgari, S., Dasgupta, N., & Stout, J. G. (2012). When do counterstereotypic ingroup members inspire versus deflate? The effect of successful professional women on young women's leadership self-concept. *Personality and Social Psychology Bulletin*, *38*(3), 370-383.

Ashcraft, C., Eger, E., & Friend, M. (2012). Girls in iT: the facts. *National Center for Women & IT. Boulder, CO*.

Babad, E. (2009). *The social psychology of the classroom*. New York: Routledge.

Beasley, M. A., & Fischer, M. J. (2012). Why they leave: The impact of stereotype threat on the attrition of women and minorities from science, math and engineering majors. *Social Psychology of Education*, *15*(4), 427-448.

Beede, D. N., Julian, T. A., Langdon, D., McKittrick, G., Khan, B., & Doms, M. E. (2011). Women in STEM: A gender gap to innovation. *Economics and Statistics Administration Issue Brief*, (04-11).

Behrens, T. E., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nature neuroscience*, *10*(9), 1214.

Benbow, C. P., & Stanley, J. C. (1982). Consequences in high school and college of sex differences in mathematical reasoning ability: A longitudinal perspective. *American Educational Research Journal*, *19*(4), 598-622.

Ben-Zeev, A., Dennehy, T.C., Sackman, R., Olide, A., & Berger, C.C. (2011). Flirting with threat: Socia identity and the perils of the female communiality prescription. *Journal of Experimental Social Psychology, 47*(1), 1308-1311

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends in cognitive sciences*, *8*(12), 539-546.

Bosson, J. K., Haymovitz, E. L., & Pinel, E. C. (2004). When saying and doing diverge: The effects of stereotype threat on self-reported versus non-verbal anxiety. *Journal of Experimental Social Psychology*, *40*(2), 247-255.

Branch, E.H. & Alegria, S. (2016). *Pathways, potholes and the persistence of women in science.* Lanham, Maryland: Lexington Books

Bridgeman, B., & Wendler, C. (1991). Gender differences in predictors of college mathematics performance and in college mathematics course grades. *Journal of Educational Psychology*, *83*(2), 275.

Brophy, J. E., & Good, T. L. (1970). Teachers' communication of differential expectations for children's classroom performance: Some behavioral data. *Journal of educational psychology*, *61*(5), 365.

Brown, R. P. & Josephs, R.A. (2000). The importance of importance: The mathematics identification questionnaire. *Unpublished manuscript: The University of Texas at Austin*

Brown, J. W., & Braver, T. S. (2005). Learned predictions of error likelihood in the anterior cingulate cortex. *Science*, *307*(5712), 1118-1121.

Brozovich, F., & Heimberg, R. G. (2008). An analysis of post-event processing in social anxiety disorder. *Clinical psychology review*, *28*(6), 891-903.

Cacioppo, J. T., Berntson, G. G., & Decety, J. (2010). Social neuroscience and its relationship to social psychology. *Social cognition*, *28*(6), 675.

Cadinu, M., Maass, A., Rosabianca, A., & Kiesner, J. (2005). Why do women underperform under stereotype threat? Evidence for the role of negative thinking. *Psychological science*, *16*(7), 572-578.

Cech, E., Rubineau, B., Silbey, S., & Seron, C. (2011). Professional role confidence and gendered persistence in engineering. *American Sociological Review*, *76*(5), 641-666.

Ceci, S. J., & Williams, W. M. (2010). Sex differences in math-intensive fields. *Current Directions in Psychological Science*, 0963721410383241.

Chase, H. W., Swainson, R., Durham, L., Benham, L., & Cools, R. (2011). Feedback-related negativity codes prediction error but not behavioral adjustment during probabilistic reversal learning. *Journal of Cognitive Neuroscience*, *23*(4), 936-946.

Cheryan, S. (2012). Understanding the paradox in math-related fields: Why do some gender gaps remain while others do not?. *Sex Roles*, *66*(3-4), 184-190.

Cheryan, S., Plaut, V. C., Davies, P. G., & Steele, C. M. (2009). Ambient belonging: How stereotypical cues impact gender participation in computer science. *Journal of personality and social psychology*, *97*(6), 1045.

Clayson, P. E., Clawson, A., & Larson, M. J. (2011). Sex differences in electrophysiological indices of conflict monitoring. *Biological psychology*, *87*(2), 282-289.

Cody, M. W., & Teachman, B. A. (2010). Post-event processing and memory bias for performance feedback in social anxiety. *Journal of Anxiety Disorders*, *24*(5), 468-479.

Cragg, L., & Gilmore, C. (2014). Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends in Neuroscience and Education*, *3*(2), 63-68.

Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological review*, *96*(4), 608.

Dasgupta, N. (2011). Ingroup experts and peers as social vaccines who inoculate the self-concept: The stereotype inoculation model. Psychological Inquiry, 22(4), 231-246.

Dasgupta, N., & Asgari, S. (2004). Seeing is believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. *Journal of Experimental Social Psychology*, *40*(5), 642-658.

Dasgupta, N., Scircle, M. M., & Hunsinger, M. (2015). Female peers in small work groups enhance women's motivation, verbal participation, and career aspirations in engineering. *Proceedings of the National Academy of Sciences*, *112*(16), 4988-4993.

Dasgupta, N., & Stout, J. G. (2014). Girls and Women in Science, Technology, Engineering, and Mathematics STEMing the Tide and Broadening Participation in STEM Careers. *Policy Insights from the Behavioral and Brain Sciences*, *1*(1), 21-29.

Deemer, E. D., Thoman, D. B., Chase, J. P., & Smith, J. L. (2014). Feeling the Threat Stereotype Threat as a Contextual Barrier to Women's Science Career Choice Intentions. *Journal of Career Development*, *41*(2), 141-158.

Derks, B., Inzlicht, M., & Kang, S. (2008). The neuroscience of stigma and stereotype threat. *Group Processes & Intergroup Relations*, *11*(2), 163-181.

Devos, T., & Banaji, M. R. (2003). Implicit self and identity. *Annals of the New York Academy of Sciences*, *1001*(1), 177-211.

Dhar, M., Wiersema, J. R., & Pourtois, G. (2011). Cascade of neural events leading from error commission to subsequent awareness revealed using EEG source imaging. *PLoS One*, *6*(5), e19578.

Endrass, T., Klawohn, J., Schuster, F., & Kathmann, N. (2008). Overactive performance monitoring in obsessive-compulsive disorder: ERP evidence from correct and erroneous reactions. *Neuropsychologia*, *46*(7), 1877-1887.

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: attentional control theory. *Emotion*, *7*(2), 336.

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. *Electroencephalography and clinical neurophysiology*, *78*(6), 447-455.

Ferreira, M. (2003). Gender issues related to graduate student attrition in two science departments. *International Journal of Science Education*, *25*(8), 969-989.

Ferriman, K., Lubinski, D., & Benbow, C. P. (2009). Work preferences, life values, and personal views of top math/science graduate students and the profoundly gifted: Developmental changes and gender differences during emerging adulthood and parenthood. *Journal of personality and social psychology*, *97*(3), 517.

Fogliati, V. J., & Bussey, K. (2013). Stereotype threat reduces motivation to improve effects of stereotype threat and feedback on women's intentions to improve mathematical ability. *Psychology of Women Quarterly*, 0361684313480045.

Forbes, C. E., & Leitner, J. B. (2014). Stereotype threat engenders neural attentional bias toward negative feedback to undermine performance. *Biological psychology*, *102*, 98-107.

Forbes, C. E., Schmader, T., & Allen, J. J. (2008). The role of devaluing and discounting in performance monitoring: A neurophysiological study of minorities under threat. *Social cognitive and affective neuroscience*, *3*(3), 253-261.

Galdi, S., Cadinu, M., & Tomasetto, C. (2014). The roots of stereotype threat: when automatic associations disrupt girls' math performance. *Child Development*, *85*(1), 250-263.

Good, J. J., Woodzicka, J. A., & Wingfield, L. C. (2010). The effects of gender stereotypic and counter-stereotypic textbook images on science performance. *The Journal of social psychology*, *150*(2), 132-147.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological review*, *102*(1), 4.

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological review*, *109*(1), 3.

Glass, J. L., Sassler, S., Levitte, Y., & Michelmore, K. M. (2014). What's so special about STEM? A comparison of women's retention in STEM and professional occupations. *Social forces*, *92*(2), 723-756.

Gonzales, P. M., Blanton, H., & Williams, K. J. (2002). The effects of stereotype threat and double-minority status on the test performance of Latino women. *Personality and social psychology bulletin*, *28*(5), 659-670.

Gudjonsson, G. H., & Sigurdsson, J. F. (2003). The relationship of compliance with coping strategies and self-esteem. *European Journal of Psychological Assessment*, *19*(2), 117.

Gudjonsson, G. H., Sigurdsson, J. F., Brynjólfsdóttir, B., & Hreinsdottir, H. (2002). The relationship of compliance with anxiety, self-esteem, paranoid thinking and anger. *Psychology, Crime and Law*, *8*(2), 145-153.

Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, gender, and math. *SCIENCE-NEW YORK THEN WASHINGTON-*, *320*(5880), 1164.

Gunderson, E. A., Ramirez, G., Levine, S. C., & Beilock, S. L. (2012). The role of parents and teachers in the development of gender-related math attitudes. *Sex Roles*, *66*(3-4), 153-166.

Gu, R., Ge, Y., Jiang, Y., & Luo, Y. J. (2010). Anxiety and outcome evaluation: the good, the bad and the ambiguous. *Biological psychology*, *85*(2), 200-206.

Haan, M. D., & Thomas, K. M. (2002). Applications of ERP and fMRI techniques to developmental science. *Developmental Science*, *5*(3), 335-343.

Hajcak, G., McDonald, N., & Simons, R. F. (2004). Error-related psychophysiology and negative affect. *Brain and cognition*, *56*(2), 189-197.

Halberda, J., Mazzocco, M., & Feigenson, L. (2008). Individual differences in nonverbal number acuity predict maths achievement. *Nature*, *455*, 665-668

Hart, A. J., Whalen, P. J., Shin, L. M., McInerney, S. C., Fischer, H., & Rauch, S. L. (2000). Differential response in the human amygdala to racial outgroup vs ingroup face stimuli. *Neuroreport*, *11*(11), 2351-2354.

Hauser, T. U., Iannaccone, R., Stämpfli, P., Drechsler, R., Brandeis, D., Walitza, S., & Brem, S. (2014). The feedback-related negativity (FRN) revisited: new insights into the localization, meaning and network organization. *Neuroimage*, *84*, 159-168.

Heldmann, M., Rüsseler, J., & Münte, T. F. (2008). Internal and external information in error processing. *BMC neuroscience*, *9*(1), 33.

Henry, J.D., & Crawford, J.R. (2005). The short-form version of the depression anxiety stress scales (DASS-21): Sonstruct validity and normative data in a large non-clinical sample. *British Journal of Clinical Psychology, 44*(2), 227-239

Herrmann, M. J., Römmler, J., Ehlis, A. C., Heidrich, A., & Fallgatter, A. J. (2004). Source localization (LORETA) of the error-related-negativity (ERN/Ne) and positivity (Pe). *Cognitive Brain Research*, *20*(2), 294-299.

Hopko, D.R., Mahadevan, R., Bare, R.L., & Hunt, M.K. (2003). The abbreviated math anxiety scale (AMAS): Construction, validity, and reliability. *Assessment, 10*(2), 178-182

Hughes, G., & Yeung, N. (2011). Dissociable correlates of response conflict and error awareness in error-related brain activity. *Neuropsychologia*, *49*(3), 405-415.

Huguet, P., & Régner, I. (2009). Counter-stereotypic beliefs in math do not protect school girls from stereotype threat. *Journal of Experimental Social Psychology*, *45*(4), 1024-1027.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science*, *321*(5888), 494-495

Hyde, J. S., & Mertz, J. E. (2009). Gender, culture, and mathematics performance. *Proceedings of the National Academy of Sciences*, *106*(22), 8801-8807.

Ibarra, H., Carter, N. M., & Silva, C. (2010). Why men still get more promotions than women. *Harvard Business Review*, *88*(9), 80-85.

Jackson, C. D., & Leffingwell, R. J. (1999). The role of instructors in creating math anxiety in students from kindergarten through college. *The Mathematics Teacher*, *92*(7), 583-586.

Johns, M., Inzlicht, M., & Schmader, T. (2008). Stereotype threat and executive resource depletion: examining the influence of emotion regulation. *Journal of Experimental Psychology: General*, *137*(4), 691.

Kashdan, T. B., & Roberts, J. E. (2007). Social anxiety, depressive symptoms, and post-event rumination: Affective consequences and social contextual influences. *Journal of Anxiety Disorders*, *21*(3), 284-301.

Keller, J. (2007). Stereotype threat in classroom settings: The interactive effect of domain identification, task difficulty and stereotype threat on female students' maths performance. *British Journal of Educational Psychology*, *77*(2), 323-338.

Krendl, A. C., Macrae, C. N., Kelley, W. M., Fugelsang, J. A., & Heatherton, T. F. (2006). The good, the bad, and the ugly: An fMRI investigation of the functional anatomic correlates of stigma. *Social Neuroscience,*, *1*(1), 5-15.

Ladouceur, C. D., Dahl, R. E., Birmaher, B., Axelson, D. A., & Ryan, N. D. (2006). Increased error-related negativity (ERN) in childhood anxiety disorders: ERP and source localization. *Journal of Child Psychology and Psychiatry*, *47*(10), 1073-1082.

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: a meta-analysis. *Psychological bulletin*, *136*(6), 1123.

Luu, P., Tucker, D. M., & Makeig, S. (2004). Frontal midline theta and the error-related negativity: neurophysiological mechanisms of action regulation. *Clinical Neurophysiology*, *115*(8), 1821-1835.

Magno, E., Foxe, J. J., Molholm, S., Robertson, I. H., & Garavan, H. (2006). The anterior cingulate and error avoidance. *Journal of Neuroscience*, *26*(18), 4769-4773.

Mangels, J. A., Good, C., Whiteman, R. C., Maniscalco, B., & Dweck, C. S. (2012). Emotion blocks the path to learning under stereotype threat. *Social Cognitive and Affective Neuroscience*, *7*(2), 230-241.

Major, B., Spencer, S., Schmader, T., Wolfe, C., & Crocker, J. (1998). Coping with negative stereotypes about intellectual performance: The role of psychological disengagement. *Personality and social psychology bulletin*, *24*(1), 34-50.

Marx, D. M., & Roman, J. S. (2002). Female role models: Protecting women's math test performance. *Personality and Social Psychology Bulletin*, *28*(9), 1183-1193.

Massar, S. A. A., Rossi, V., Schutter, D. J. L. G., & Kenemans, J. L. (2012). Baseline EEG theta/beta ratio and punishment sensitivity as biomarkers for feedback-related negativity (FRN) and risk-taking. *Clinical Neurophysiology*, *123*(10), 1958-1965.

Matchock, R. L., & Mordkoff, J. T. (2009). Chronotype and time-of-day influences on the alerting, orienting, and executive components of attention. *Experimental brain research*, *192*(2), 189-198.

McArdle, E. (2008). The freedom to say 'no.' The Boston Globe. Retrieved from http://www.boston.com/bostonglobe/ideas/articles/2008/05/18/the_freedom_to_say_no/

McIntyre, R. B., Lord, C. G., Gresky, D. M., Ten Eyck, L. L., Frye, G. J., & Bond Jr, C. F. (2005). A social impact trend in the effects of role models on alleviating women's mathematics stereotype threat. *Current Research in Social Psychology*, *10*(9), 116-36.

Medin, D. L., & Lee, C. D. (2012). Diversity makes better science. *Association for Psychological Science Observer*.

Mizala, A., Martínez, F., & Martínez, S. (2015). Pre-service elementary school teachers' expectations about student performance: How their beliefs are affected by their mathematics anxiety and student's gender. *Teaching and Teacher Education*, *50*, 70-78.

Morgan, J., & Banerjee, R. (2008). Post-event processing and autobiographical memory in social anxiety: The influence of negative feedback and rumination. *Journal of Anxiety Disorders*, *22*(7), 1190-1204.

Moser, J. S., Moran, T. P., & Jendrusina, A. A. (2012). Parsing relationships between dimensions of anxiety and action monitoring brain potentials in female undergraduates. *Psychophysiology*, *49*(1), 3-10.

Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat how situational cues affect women in math, science, and engineering settings. *Psychological Science*, *18*(10), 879-885.

National Science Foundation, National Center for Science and Engineering Statistics. 2015. *Women, Minorities, and Persons with Disabilities in Science and Engineering: 2015*. Special Report NSF 15-311. Arlington, VA. Available at http://www.nsf.gov/statistics/wmpd/.

Neuville, E., & Croizet, J. C. (2007). Can salience of gender identity impair math performance among 7–8 years old girls? The moderating role of task difficulty. *European Journal of Psychology of Education*, *22*(3), 307-316.

Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P., & Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: Evidence from an antisaccade task. *Psychophysiology*, *38*(5), 752-760.

Nussbaum, A. D., & Steele, C. M. (2007). Situational disengagement and persistence in the face of adversity. *Journal of experimental social psychology*, *43*(1), 127-134.

Orr, J. M., & Carrasco, M. (2011). The role of the error positivity in the conscious perception of errors. *Journal of Neuroscience*, *31*(16), 5891-5892.

Osborne, J. W. (2007). Linking stereotype threat and anxiety. *Educational Psychology*, *27*(1), 135-154.

Osborne, J. W., & Walker, C. (2006). Stereotype threat, identification with academics, and withdrawal from school: Why the most successful students of colour might be most likely to withdraw. *Educational Psychology*, *26*(4), 563-577.

Overbeek, T. J., Nieuwenhuis, S., & Ridderinkhof, K. R. (2005). Dissociable components of error processing: On the functional significance of the Pe vis-à-vis the ERN/Ne. *Journal of Psychophysiology*, *19*(4), 319-329.

Pham, N. D., & Triantis, A. J. (2015). Reaching the full potential of STEM for women and the US economy. *Washington, DC: US Chamber of Commerce Foundation*.

Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of cognitive neuroscience*, *12*(5), 729-738.

Räty, H., Vänskä, J., Kasanen, K., & Kärkkäinen, R. (2002). Parents' explanations of their child's performance in mathematics and reading: A replication and extension of Yee and Eccles. *Sex roles*, *46*(3-4), 121-128.

Ridderinkhof, K. R., Ramautar, J. R., & Wijnen, J. G. (2009). To PE or not to PE: A P3-like ERP component reflecting the processing of response errors. *Psychophysiology*, *46*(3), 531-538.

Rosenbloom, J. L., Ash, R. A., Dupont, B., & Coder, L. (2008). Why are there so few women in information technology? Assessing the role of personality in career choices. *Journal of Economic Psychology*, *29*(4), 543-554.

Rosenthal, H. E., & Crisp, R. J. (2006). Reducing stereotype threat by blurring intergroup boundaries. *Personality and Social Psychology Bulletin*, *32*(4), 501-511.

Rydell, R. J., Van Loo, K. J., & Boucher, K. L. (2013). Stereotype Threat and Executive Functions Which Functions Mediate Different Threat-Related Outcomes?. *Personality and Social Psychology Bulletin*, 0146167213513475.

Sachs, G., Anderer, P., Margreiter, N., Semlitsch, H., Saletu, B., & Katschnig, H. (2004). P300 event-related potentials and cognitive function in social phobia. *Psychiatry Research: Neuroimaging*, *131*(3), 249-26S1.

Sadler, P. M., Sonnert, G., Hazari, Z., & Tai, R. (2012). Stability and volatility of STEM career interest in high school: A gender study. *Science Education*, *96*(3), 411-427.

Schaefer, A., Buratto, L. G., Goto, N., & Brotherhood, E. V. (2016). The feedback-related negativity and the P300 brain potential are sensitive to price expectation violations in a virtual shopping task. *PloS one*, *11*(9), e0163150.

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of personality and social psychology*, *85*(3), 440.

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological review*, *115*(2), 336.

Sekaquaptewa, D., & Thompson, M. (2003). Solo status, stereotype threat, and performance expectancies: Their effects on women's performance. *Journal of Experimental Social Psychology*, *39*(1), 68-74.

Shapiro, J. R., & Williams, A. M. (2012). The role of stereotype threats in undermining girls' and women's performance and interest in STEM fields. *Sex Roles*, *66*(3-4), 175-183.

Shields, S. (1975). Functionalism, Darwinism, and the psychology of women. *American Psychologist*, *30*(7), 739.

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological science*, *10*(1), 80-83.

Silvetti, M., Seurinck, R., & Verguts, T. (2013). Value and prediction error estimation account for volatility effects in ACC: a model-based fMRI study. *cortex*, *49*(6), 1627-1635.

Spears, R., Doosje, B., & Ellemers, N. (1997). Self-stereotyping in the face of threats to group status and distinctiveness: The role of group identification. *Personality and Social Psychology Bulletin*, *23*(5), 538-553.

Stahl, J. (2010). Error detection and the use of internal and external error indicators: An investigation of the first-indicator hypothesis. *International Journal of Psychophysiology*, *77*(1), 43-52.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American psychologist*, *52*(6), 613.

Stone, J., & McWhinnie, C. (2008). Evidence that blatant versus subtle stereotype threat cues impact performance through dual processes. *Journal of Experimental Social Psychology*, *44*(2), 445-452.

Stout, J. G., Dasgupta, N., Hunsinger, M., & McManus, M. A. (2011). STEMing the tide: using in-group experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM). *Journal of personality and social psychology*, *100*(2), 255.

Summers, L.H. (2005, January 14). Remarks at NBER conference on diversifying the science and engineering workforce. Retreived from: http://www.harvard.edu/president/speeches/summers_2005/nber.php

Talmi, D., Fuentemilla, L., Litvak, V., Duzel, E., & Dolan, R. J. (2012). An MEG signature corresponding to an axiomatic model of reward prediction error. *Neuroimage*, *59*(1), 635-645.

Tanovic, E., Hajcak, G., & Sanislow, C. A. (2017). Rumination is associated with diminished performance monitoring. *Emotion*, *17*(6), 953.

Tiedemann, J. (2000). Parents' gender stereotypes and teachers' beliefs as predictors of children's concept of their mathematical ability in elementary school. *Journal of Educational psychology*, *92*(1), 144.

Ullsperger, M., Harsay, H. A., Wessel, J. R., & Ridderinkhof, K. R. (2010). Conscious perception of errors and its relation to the anterior insula. *Brain Structure and Function*, *214*(5-6), 629-643.

UMasss Office of Institutional Research (2017). *Race/Ethnicity of undergraduate students (U.S. Citizens).* Retrieved from https://www.umass.edu/oir/students/diversity

U.S. Census Bureau (2016). American Community Survey 1-year estimates. Retrieved from Census Reporter Profile page for Springfield, MA <https://censusreporter.org/profiles/16000US2567000-springfield-ma/>

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin*, *140*(4), 1174.

Walker, C. O., Greene, B. A., & Mansell, R. A. (2006). Identification with academics, intrinsic/extrinsic motivation, and self-efficacy as predictors of cognitive engagement. *Learning and individual differences*, *16*(1), 1-12.

Wang, M. T., Eccles, J. S., & Kenny, S. (2013). Not lack of ability but more choice individual and gender differences in choice of careers in science, technology, engineering, and mathematics. *Psychological Science*, *24*(5), 770-775.

Weeks, J. W., Heimberg, R. G., Rodebaugh, T. L., & Norton, P. J. (2008). Exploring the relationship between fear of positive evaluation and social anxiety. *Journal of Anxiety Disorders*, *22*(3), 386-400.

Weinberg, A., Olvet, D. M., & Hajcak, G. (2010). Increased error-related brain activity in generalized anxiety disorder. *Biological psychology*, *85*(3), 472-480.

Wheeler, M. E., & Fiske, S. T. (2005). Controlling racial prejudice social-cognitive goals affect amygdala and stereotype activation. *Psychological Science*, *16*(1), 56-63.

Woodcock, A., Hernandez, P. R., Estrada, M., & Schultz, P. (2012). The consequences of chronic stereotype threat: domain disidentification and abandonment. *Journal of personality and social psychology*, *103*(4), 635.

Wraga, M., Duncan, L., Jacobs, E. C., Helt, M., & Church, J. (2006). Stereotype susceptibility narrows the gender gap in imagined self-rotation performance. *Psychonomic Bulletin & Review*, *13*(5), 813-819.

Wraga, M., Helt, M., Jacobs, E., & Sullivan, K. (2006). Neural basis of stereotype-induced shifts in women's mental rotation performance. *Social cognitive and affective neuroscience*.

Xiao, Z., Wang, J., Zhang, M., Li, H., Tang, Y., Wang, Y., ... & Fromson, J. A. (2011). Error-related negativity abnormalities in generalized anxiety disorder and obsessive–compulsive disorder. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *35*(1), 265-272.

Xu, Y. J. (2008). Gender disparity in STEM disciplines: A study of faculty attrition and turnover intentions. Research in Higher Education, 49(7), 607-624.

Xu, Q., Shen, Q., Chen, P., Ma, Q., Sun, D., & Pan, Y. (2011). How an uncertain cue modulates subsequent monetary outcome evaluation: an ERP study. *Neuroscience Letters*, *505*(2), 200-204.

Yee, D. K., & Eccles, J. S. (1988). Parent perceptions and attributions for children's math achievement. *Sex Roles*, *19*(5-6), 317-333.